

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:55:28 ; Search time 134.581 Seconds
(without alignments)
1613.049 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTTTSGSPMPDGTGVL.....SIDGDMCSGLVYTGVDQCA 676

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 segs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3528	100.0	676	9	US-09-840-762A-2
2	3528	100.0	676	16	US-10-691-383-2
3	130.5	3.7	3192	14	US-10-132-134-10
4	127.5	3.6	650	15	US-10-369-493-18669
5	123.5	3.5	3798	13	US-10-014-717-6
6	123	3.5	1939	12	US-10-152-886-43
7	122	3.5	22	9	US-09-840-762A-9
8	122	3.5	22	16	US-10-691-383-9
9	120.5	3.4	6310	12	US-10-282-122A-67793
10	117	3.3	888	14	US-10-156-761-8791
11	117	3.3	1387	16	US-10-437-963-167961
12	116	3.3	5245	14	US-10-329-079-11
13	114.5	3.2	912	9	US-09-746-491-49
14	114	3.2	2365	14	US-10-156-761-14890
15	113.5	3.2	912	9	US-09-905-983-2

16	113.5	3.2	3472	13	US-10-027-806-4	Sequence 4, Appli
17	113.5	3.2	3472	13	US-10-034-823-4	Sequence 4, Appli
18	113.5	3.2	3472	14	US-10-027-801-4	Sequence 4, Appli
19	113.5	3.2	3472	14	US-10-029-120-4	Sequence 4, Appli
20	113	3.2	4999	9	US-09-976-059-15	Sequence 15, Appli
21	112.5	3.2	730	16	US-10-437-963-122066	Sequence 122066,
22	111.5	3.2	1385	14	US-10-169-297-46	Sequence 46, Appli
23	111.5	3.2	1387	14	US-10-169-297-14	Sequence 14, Appli
24	111	3.1	391	16	US-10-437-963-132036	Sequence 132036,
25	110	3.1	21	16	US-10-691-383-11	Sequence 11, Appli
26	109.5	3.1	453	14	US-10-156-761-8288	Sequence 8288, Ap
27	109.5	3.1	1965	15	US-10-369-493-2646	Sequence 2646, Ap
28	109	3.1	611	15	US-10-369-493-12809	Sequence 12809, A
29	108.5	3.1	165	14	US-10-080-170-629	Sequence 629, App
30	108.5	3.1	165	16	US-10-080-170-629	Sequence 629, App
31	108.5	3.1	1439	12	US-10-282-122A-44567	Sequence 44567, A
32	108	3.1	19	9	US-09-840-762A-10	Sequence 10, Appli
33	108	3.1	19	16	US-10-691-383-10	Sequence 10, Appli
34	107.5	3.0	637	12	US-10-282-122A-69863	Sequence 69863, A
35	107.5	3.0	2394	14	US-10-329-079-47	Sequence 47, Appli
36	107	3.0	1273	12	US-10-363-616-325	Sequence 325, App
37	107	3.0	2273	12	US-10-282-122A-66115	Sequence 66115, A
38	107	3.0	2435	12	US-10-282-122A-47453	Sequence 47453, A
39	106.5	3.0	947	12	US-10-282-122A-63665	Sequence 63665, A
40	106	3.0	303	12	US-10-282-122A-69701	Sequence 69701, A
41	106	3.0	426	15	US-10-369-493-1118	Sequence 1118, Ap
42	106	3.0	1084	16	US-10-437-963-152384	Sequence 152384, A
43	106	3.0	3705	12	US-10-282-122A-77944	Sequence 77944, A
44	105.5	3.0	451	14	US-10-156-761-10605	Sequence 10605, A
45	105.5	3.0	632	15	US-10-369-493-15431	Sequence 15431, A

ALIGNMENTS

RESULT 1

US-09-840-762A-2
; Sequence 2, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-08710005
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
; US-09-840-762A-2

Query Match	100.0%;	Score 3528;	DB 9;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	676;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	MLCHAADTTTSGSPMPDGTGVLRLTSEORAKGWRQLEGEKSLGFPHPSETPYIKYLEGSET 60		
Db	1	MLCHAADTTTSGSPMPDGTGVLRLTSEORAKGWRQLEGEKSLGFPHPSETPYIKYLEGSET 60		
Qy	61	WKVKUPTDGISASKILGKIMARVATATALLAVLAAPCLAFDEVTASGVFPEEHKHGTGEG 120		
Db	61	WKVKUPTDGISASKILGKIMARVATATALLAVLAAPCLAFDEVTASGVFPEEHKHGTGEG 120		
Qy	121	RHLQTCNSDDADPTAPNRDNVAFASRRDARRRDGTTGCOITNGTDLATMPHKS 180		
Db	121	RHLQTCNSDDADPTAPNRDNVAFASRRDARRRDGTTGCOITNGTDLATMPHKS 180		

181 LPHELGQVTTADDFAILEDCLNGDFSI CEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT 240
181 LPHELGQVTTADDFAILEDCLNGDFSI CEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT 240
241 TTPVPTLSSPELAQAELLYMALARVDFPMQYGTDEITTTAAANLAGMGGFPNLDVAVS 300
241 TTPVPTLSSPELAQAELLYMALARVDFPMQYGTDEITTTAAANLAGMGGFPNLDVAVS 300
301 IGSDGTVDPSQLPRATFVGVEGTFVPSQLLVNSFTTDAITVBPQKQTFAPDLNMYWDFD 360
301 IGSDGTVDPSQLPRATFVGVEGTFVPSQLLVNSFTTDAITVBPQKQTFAPDLNMYWDFD 360
361 EWLNIQNGGPPAGPEELDEELRFRNARDLARVSFVDNINTEAYRGSLLIILELGAFSRPG 420
361 EWLNIQNGGPPAGPEELDEELRFRNARDLARVSFVDNINTEAYRGSLLIILELGAFSRPG 420
421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGTLHNTI 480
421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGTLHNTI 480
481 AGDLDDADFDLSLENDLLKRVABINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQ 540
481 AGDLDDADFDLSLENDLLKRVABINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQ 540
541 NGAPATVVKALIGLDRGCEFPNVPFSDGDLLEINPEGACLTVEGINKLVANVAFGRQ 600
541 NGAPATVVKALIGLDRGCEFPNVPFSDGDLLEINPEGACLTVEGINKLVANVAFGRQ 600
601 MLGIHYRFDGIQGLLGETITVTRLHQELMTFAEEATFEPRFTGEVIKLPQDGTFSIDG 660
601 MLGIHYRFDGIQGLLGETITVTRLHQELMTFAEEATFEPRFTGEVIKLPQDGTFSIDG 660
661 DMCSGLVYTGADCOA 676
661 DMCSGLVYTGADCOA 676

RESULT 2

US-10-691-383-2
; Sequence 2, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-0871110US
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/10/691,383
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 1998-09-10
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-10-691-383-2

Query Match 100.0%; Score 3528; DB 16; Length 676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLCHAADTRGSPMPDGTGVLRLTSEQRAKGRWRLQEGEKSGLGPHSPETPIYKYLEGSET 60
1 MLCHAADTRGSPMPDGTGVLRLTSEQRAKGRWRLQEGEKSGLGPHSPETPIYKYLEGSET 60
61 WKVKLPDGTGSAKILGKIMARVRIATALAVLAAPCLAFDEVTASGVPEEHKHTGEG 120
61 WKVKLPDGTGSAKILGKIMARVRIATALAVLAAPCLAFDEVTASGVPEEHKHTGEG 120

121 RHLQTCNDDALDPTAPNRDNDVAFASRRDARRERDGTGTVCQITNGETDLATMPHKS 180
121 RHLQTCNDDALDPTAPNRDNDVAFASRRDARRERDGTGTVCQITNGETDLATMPHKS 180
181 LPHELGQVTTADDFAILEDCLNGDFSI CEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT 240
181 LPHELGQVTTADDFAILEDCLNGDFSI CEDVPAGDPAGRLVNPTAAFAIDISGPAFSAT 240
241 TTPVPTLSSPELAQAELLYMALARVDFPMQYGTDEITTTAAANLAGMGGFPNLDVAVS 300
241 TTPVPTLSSPELAQAELLYMALARVDFPMQYGTDEITTTAAANLAGMGGFPNLDVAVS 300
301 IGSDGTVDPSQLPRATFVGVEGTFVPSQLLVNSFTTDAITVBPQKQTFAPDLNMYWDFD 360
301 IGSDGTVDPSQLPRATFVGVEGTFVPSQLLVNSFTTDAITVBPQKQTFAPDLNMYWDFD 360
361 EWLNIQNGGPPAGPEELDEELRFRNARDLARVSFVDNINTEAYRGSLLIILELGAFSRPG 420
361 EWLNIQNGGPPAGPEELDEELRFRNARDLARVSFVDNINTEAYRGSLLIILELGAFSRPG 420
421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGTLHNTI 480
421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGTLHNTI 480
481 AGDLDDADFDLSLENDLLKRVABINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQ 540
481 AGDLDDADFDLSLENDLLKRVABINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQ 540
541 NGAPATVVKALIGLDRGCEFPNVPFSDGDLLEINPEGACLTVEGINKLVANVAFGRQ 600
541 NGAPATVVKALIGLDRGCEFPNVPFSDGDLLEINPEGACLTVEGINKLVANVAFGRQ 600
601 MLGIHYRFDGIQGLLGETITVTRLHQELMTFAEEATFEPRFTGEVIKLPQDGTFSIDG 660
601 MLGIHYRFDGIQGLLGETITVTRLHQELMTFAEEATFEPRFTGEVIKLPQDGTFSIDG 660
661 DMCSGLVYTGADCOA 676
661 DMCSGLVYTGADCOA 676

RESULT 3

US-10-132-134-10
; Sequence 10, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 3192
; TYPE: PRT
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-10

Query Match 3.7%; Score 130.5; DB 14; Length 3192;
Best Local Similarity 19.0%; Pred. No. 0.053;
Matches 150; Conservative 84; Mismatches 254; Indels 303; Gaps 38;

5 AADTTRGSPMPDGTGVLRLTSEQRAKGRWRLQEGEKSGLGPHSPETPIYKYLEGSET 45
2228 AADVPSAPAEADASMPQMSAPATLDRDDFVAALGTAGLDARPYARSVEGVTELDARLL 2287
46 -----PS--ETPIYKYLEGSETWKVKLPDGTGSI-----ASKILGKIMARVRI 86
2288 VRVAEPAMCQDPKHQVH-LPAWALVGL-TQGVQAWGRADAADVVRVGVQGEQWERT- 2344

Db 2814 LEILPFGSGFDMABRIYRDSPIARYSGNIVRGVVEAARVAVVAPSGMFSEILLEIGAGTGATT 2873
 Qy 242 IPPVPTLSSPELAQLAELYMALA-----RDVPFMOYGTDEITTTTAAANLAGMGGFNP 295
 Db 2874 AAVLPVLLPORTEYHFTDVSPFLFARAEQRFRDYFFLYKGLVDQSPAGQGYAHORFDV 2933
 Qy 296 LDAVSGSDGTVDPFSQLFRATFVGTVGTFVSPVQLLVNSFTIDAITVEPKQETPAP-DLN 354
 Db 2934 IVAANV-IHATRD-----IRAT-----AKELL-----SLLAFGGLL 2963
 Qy 355 YMDV---FDEWLNTQNGPPAGPELDELRF---IRNARD---LARVSFVNI----- 399
 Db 2964 VLVEGTGHPWFDTITT-GLIEGWOKYEDDLRIDHPLLPARTCDVLRVGFADAVSLPGD 3022
 Qy 400 -NTEAYGSLILLELGFASRPGINGPFDSDRQAGFVNFCTSHYFRLLIGAEAAQASCY 458
 Db 3023 GSPAGILGQHVLS-----RAGTIAGAACDSGES-----ATESPAARAVR 3063
 Qy 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
 Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGLRTGGAFYKALAGDLLLPEDTGQVV 3123
 Qy 491 SLENDELLKRVAEINAAQNNEVTVLLPOAIQVGSPTHPSPSGHAT-----QNG 542
 Db 3124 AEVQGLRLPOLEASAFAPRDPREEWLYALEWQRKDPFPEAPAAASSSSAGAWLVLMDOGG 3183
 Qy 543 AFATVLKALJGLDRGGEC 560
 Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 6
 US-10-152-886-43
 ; Sequence 43, Application US/10152886
 ; Publication No. US20030064491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ECOPIA BIOSCIENCES INC.
 ; APPLICANT: Farnet, Chris
 ; APPLICANT: Staffa, Alfredo
 ; APPLICANT: Zazopoulos, Emmanuel
 ; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
 ; FILE REFERENCE: 3011-3US
 ; CURRENT APPLICATION NUMBER: US/10/152,886
 ; CURRENT FILING DATE: 2002-05-21
 ; NUMBER OF SEQ ID NOS: 102
 ; SOFTWARE: Patentin version 3.0
 ; SEQ ID NO 43
 ; LENGTH: 1939
 ; TYPE: PRT
 ; ORGANISM: Amycolatopsis orientalis
 US-10-152-886-43

Query Match 3.5%; Score 123; DB 12; Length 1939;
 Best Local Similarity 22.6%; Pred. No. 0.13;
 Matches 146; Conservative 59; Mismatches 249; Indels 192; Gaps 32;
 Qy 4 HAADTRGSPMDTGVRLITSEORAK---GWRQLEGEKSLGPHSPETPIYKLGSET 60
 Db 606 HADDIYRAANL-STGADQVATDVAQPRIVTGSGLAGRLVLSGLIEATV--TGHSLGELT 662
 Qy 61 ---WKKVKLPTDGISASKILGKMARV-----RIATALA----- 91
 Db 663 ALHWGGALTREVLKAKIRKQVWATASDGGAWAATAATPSVAEGLAESEVVVIAGYNA 722
 Qy 92 ---VVLAAFLAFDEVTA-----SGVPEEHKHTGEGRLQ--- 124
 Db 723 PEQTVLSPGAEDRVARAEAGVTAARINVSHPAFSPVPAEAMTGLAIDAIDPARL 782
 Qy 125 ----TCTNSDLDLPTAPNR---RDVAFASR-RDAARRRDGTGTCQITNGE----- 170
 Db 783 DRPVVSTVTGDLVHAEDLRLDRDQVVLVPRFREAAAKVAERSDLVIEVGPRVLTGLL 842

Qy 171 -----TDLATMFKHSLPHDELQVTDADPAI---LEDICL-----NGDFSI-- 208
 Db 843 GTIAPGTPVLSIOTDSITLAPVLKVAGAAFAFGAQLSTLFDGRVVRALPADGEFSLA 902
 Qy 209 --CEDVP---AGDPAGRLVNPTAAFAIDISGPASATTPPPVPTLSPPELAQAELIYWM 263
 Db 903 SPCEAAPSGICAVLTRDRVAPAEAAAGTASESGSST---LDLLR--KLASERVELPLE 956
 Qy 264 ALARDV-PFMOYGTDEIT-----TTAAANLAGMGFPNLDVAVSIGS-DGTVDPSOLF 314
 Db 957 AVTADTHPLDLDLHLSITVQQLVNDVTRALGRPALGMPNFATVCLGELAEMIDELAOTA 1016
 Qy 315 R-----ATFVGVEYTPFVSQQLLVNSFTIDAITVEPKQETFAFDNLVWVDFEWLNTQNG 368
 Db 1017 KPADSNQAEVAGV--GPW-----VRPFAVEV-VAKP-----SPDLATGISTAEWTAAPA 1065
 Qy 369 GPPAGPEELDEELFRIRNARDLARVSFVDNINTEAYRGSLLLELGFASRPGINGPFDIS 428
 Db 1066 GHP-----LAEPRLR-----AALATAGVGDG-----VLLCLNADSASGDVGLFLDA 1105
 Qy 429 DR-----QAGFVNEGTSHYFEL-----IGAAELAQRASCYQKQVHR 465
 Db 1106 GRAVLAAPNGTRFVVQVHGLGASGLAKTLRLDPSATTIVDLADLGP----- 1153
 Qy 466 FARPEALGGTLHNTIAGDLDADFDISLENDELLKRVAEINAAQNP 511
 Db 1154 -VDPEALDRAV-STVTEVAATTDSEVRVYDTAGVTRVTPKLAALTP 1197

RESULT 7
 US-09-840-762A-9
 ; Sequence 9, Application US/09840762A
 ; Patent No. US20020035245A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Vreeland, Valerie
 ; APPLICANT: NG, Kwan L.
 ; TITLE OF INVENTION: The Regents of the University of California
 ; FILE REFERENCE: 023070-087100US
 ; CURRENT APPLICATION NUMBER: US/09/840,762A
 ; CURRENT FILING DATE: 2001-04-23
 ; PRIOR APPLICATION NUMBER: 09/151,189
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 22
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: conserved
 ; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473
 US-09-840-762A-9

Query Match 3.5%; Score 122; DB 9; Length 22;
 Best Local Similarity 100.0%; Pred. No. 7.9e-05;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 452 AQRASCYQKQVHRFARPEALG 473
 Db 1 AQRASCYQKQVHRFARPEALG 22

RESULT 8
 US-10-691-383-9
 ; Sequence 9, Application US/10691383
 ; Publication No. US20040110260A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Vreeland, Valerie
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
 ; FILE REFERENCE: 023070-087110US


```
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,
; OTHER INFORMATION: 1st conserved motif
US-10-691-383-9

Query Match          3.5%; Score 122; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 452 AQRASCYQKQVHRFARPEALG 473
Db 1 AQRASCYQKQVHRFARPEALG 22

RESULT 9
US-10-282-122A-67793
; Sequence 67793, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Havelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUIPA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67793
; LENGTH: 6310
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-67793

Query Match          3.4%; Score 120.5; DB 12; Length 6310;
Best Local Similarity 20.2%; Pred. No. 1.6;
Matches 156; Conservative 271; Indels 279; Gaps 36;

Qy 32 WRRQLEGEKSLGPHSPETPYIKYLEGSETWKKVKKLPTDGGISASKILKIMARVRIATALA 91
Db 86 WHAQYDAAAFGTFFBEVASLDELVAGIGTAGSAMPTWAIAGLSLLG-----VGGAAA 138
Qy 92 VW-----LAAECLAFDEVTASGVPEEHKHTGEGHLOTCINSDA----- 132
Db 139 AADNGGGSSGGSDPDTSAPATPIDLL-----VSPDGLRLTGRG-EAGTTVNIRDAAGNLI 193
Qy 133 -----LDPTAPNRNRNVAFAASRRDAARRDGTGTCQITN 168
Db 194 GSGTVGADGNFNTLNAPQINSENLDVLTDAAGNV-----APGAVTAPDATAPE----- 244
Qy 169 GETDLA-----TMFHKSLPHDELGQVTADDFAILEDCI--LNGDFSICEDVPAGDPAGR 220
Db 245 APTDLAINEQGNLTGAEFGSTV-SVRGAGGVLLGTAVAGADGQFSITLQPPQSD--GQ 301
Qy 221 LVNPTAAFAIDISGPAPSATTIIPV--PTLSSPELAAQLAEIYMWALARDVPFMQYGTDE 278
Db 302 ALBISAADAAGNISPVASITA-PDVNDPDTTAEQPTDL-----ALADGVTFTRGEPG 354
Qy 279 ITT-----TAAANLAGMGFPNLDVAVSGSDCTVDPFQSOLFRTATFVGTVETGPFVSOLLVNSF 335
Db 355 ATVOVRDAAGNLTGT-----VVGADG-----LFSLTLSPAQT-----NGE 390
Qy 336 TIDAITVEPKQETFAPDLNMYWDFD-----EWLNQNGQPPAGPELD 378
Db 391 ALDVLVDAAGNSAP-----LQFDAPDITPPDAVTNITVGADGLALSGRGEPAATVB-- 443
Qy 379 EEIRFIRNARDLARVSFVDNINTEAVRGSLLLELGAFFSPGNGPFIDSDRQAGFVNFG 438
Db 444 -----VRDANGTVIGTVGVGAN-----GTFLIDLAPAAQFGEQLSLVQTD-PSGNASVA 491
Qy 439 TSHYFRLLIGRAE-----LAQRASCYQKQVHRFARPEALGGTLLHNTIAGDLD 485
Db 492 TEYDVPLTTPADSPSNLAIDADGTTLTGTAPAGSRVEVH-----DANGTLIGSAIANAD 545
Qy 486 ADPDISI-----LENDELLKRYA-----EINAA 508
Db 546 GSFSEIENPAQANGELLDVVAIDDDGGVSSLPQAQITAPDITAPAPTETELAVSADGSVITGR 605
Qy 509 QNPNEVTVLLPQAIQVGS-----PT-----HPSYPSGHATQNGAFATVKKALIGLDRG 557
Db 606 AEPGSTVRIVAADGTGELGTAVGVGTGVFSLNLPNPQVDGEVLQ--ATAT-----DAA 655
Qy 558 GECFPNPVF--PSDDGLE-----LINFEGACLTYEIGEINKLVAVNVAFGQMLGIH 605
Db 656 GNTSPSSAVTAPDIDGVDITPPAAPTDLVLVLGAGSQLSGRGE----- 697
Qy 606 YRFDGIQGLLGETITVTRLHQLMTFAEATFEFRLTGTCEVIKLFODGTFSI 658
Db 698 -----AGSTVQVRDAAGNI-----LATGTVA---ADGFTTV 725

RESULT 10
US-10-156-761-8791
; Sequence 8791, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8791
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8791

Query Match 3.3%; Score 117; DB 14; Length 888;
Best Local Similarity 21.7%; Pred. No. 0.13;
Matches 151; Conservative 85; Mismatches 255; Indels 204; Gaps 36;
6 ADTTRGSPMDTGVLRLTSEQ-----RAKGRQLEGEKS-----LGFHPSETP 50
28 ADRASSVARPAQDVLSLQTERLITAVQASRTKTARTELDGARDETDAAVAFRRSS- 86
51 YIKYLEGSETWKVKL---PTDGISASK---ILGKIMA-----RVRIATALAVVLAAP 97
87 --SALDTSLSLQRRTRLFDEALDLSGRREALDGRTLSTGDTFELYTDTVSRGIGVLTAA- 143
98 CLAFDEVASGVPPPEEHKHTGEBR-----HLQCTNSDDALDPTA--PNRR--- 141
144 -----VHSDGRLARGGNATVSLAHITELMSREDALISGALPFSRMTA 186
142 -DNVAFASRRDAARRRDG-----TGTVCOITNG-----ETDLATMPHKS 181
187 TTRAQFOYLAIQREFRAGLDIHDLPAGAAATYTLQIDSAKWTTLGTVKESVITGRGTG 246
182 PHDELQVTDADFAILEDCLINGDFISCEVDP-----AGDPA-----GRLVNPTAF 230
247 P---GQ--ASSMPTATEPVVGGIQLGADSDGLSDLAADHADDLLGLMLG--TAATLA 299
231 DISGAPASA-----TTTPPVPTLSS-----PELAQALAEYMWALARDVPFMOYG 275
300 ALAGAVLALRARRSTLGRVSELQAHVEQLSGSWLPOLLARIQNGERVEPATLAPHGEQA 359
276 TDEITTTAAANLAGMGFPNLDAV--SIGSDGTVDPSQLFRATFV-----GVET 323
360 TDELERLAAA-IDQLGRVAGDTAVRQSLGREGTEKVFQAQIRRTQIILHLISLLD 418
324 GPFVSQLLVNSFTIDAITVEPKQFTFAPDLNMYVDFDEWLNIONGGPPAGPEELBELRF 383
419 KHEDSLLKDKIFKVDHLATVRHRA-----ENLVLSGSP----- 454
384 IRNARDLARVSFVDNINTEAYRGLLILELGA SPRPGINGPFFIDSDRQAGVNFNGTSHYF 443
455 --SRRLTAPVSI-----TDVMRGA--VAETEQTTRVKVN--LPADRRLLALAGRAVD 503
444 RLIGAELAQRASCYQ--KQVHRFARPEALGGTLH-----NTIAGDLAD-- 487
504 HLL--AELIENGTSFSPDQTFVSA TKVAGLAHVHVEDHGLGMPQDLRHANDLLAHP 561
488 -FDISLLENDE-----LLKRVAEINAAQNPNNEVTV-----LLPOAI--QVGSPT 532
562 RLDMTALGEDPRLGHFVVVARLAERHKIKVELRESVYGGTLVILVLPALLEEVA 621
533 YPSGHATONGAFAT-VLKALIGLDR-----GGECP 562
622 LKSAVAANRAVAEASRAVAGTDDFAVVGAEQMP 656

RESULT 11

US-10-437-963-167961
; Sequence 167961, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 167961
; LENGTH: 1387
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66522C.1.pep
US-10-437-963-167961

Query Match 3.3%; Score 117; DB 16; Length 1387;
Best Local Similarity 18.8%; Pred. No. 0.28;
Matches 153; Conservative 89; Mismatches 273; Indels 298; Gaps 38;

26 EORAKWRQLEGEKSLGPHSPSETPYIKYL--EGSETWKVKVLPDTGISAASKILGKI-- 80
286 QSLLEYLRKQAEYDGRG-----PVIDIVAHGDGVW-RVAVDTQGLEGNKCGKLADF 338
81 --NARVRI-----ATALAVVLA--PCLAF 101
339 VPLTNYRLERKFGIFSKLDACS FVANIYDDGNLVSIVTDCSPHATHVAGIAAHP- 394
102 DEVTASGVPPPEEHKHTGEBRHLQCTNSDDALDPTAPNRDNDVAFASRRDAARRDGTG 161
395 DEPLLNGVAP-----CAQLISCRIGDTRLSGM-----ETGTG 426
162 TV-----CQITN-----GETDLATMPHKSIPHDELQVTDADFAI--LEDCLINGDF 206
427 LVRALTAAVEHKCOLINMSYGEPTLLPDYGRFI---DLASEVVDKRIIFISAGNNGPA 483
207 SICEDVPAGDPA-----GRLVNPTAA-----FAIDIS 233
484 LNTVGAPGGTSSSIIGVGA VSPAMAAGAHCVVQAPAEQMEYTWSSRGPTADGDLGVSIS 543
234 GPAFSATTTPPVPT-----LSPPELAQALAEYMWALARDVPFMOYGTDEIT 280
544 AP---GGAVAPVPTWTLSRMLMNGTSMSSPSACGGVALLVSAKAEGLPSPTYVRKAI 600
281 TTAANLAGM-----GGFPNLD-----AVSIGSDGTVDPSQLFR 315
601 ENTAASISDVPEEKLTGHLGLQVDRAFEYAAQAKELPLVSYRISINQVG--KPTSKLRG 658
316 ATVGVGVTGPFVSQLVNSFTIDAITVEPKQFTFAPDLNMYVDFDEWLNIONGGPPAGPE 375
659 IYLRGNTCQTSWTVQ-----LDPKFHEDASNMEOQVFEELQLHS--TDSIVI 708
376 ELDEELRFRINARDLARVSFVDNINT-----EAY-----RGLLILLGLAFSPRGIN 422
709 KIPEYIMVTNNGRTFNILVNPVNISSGLHYEYVYGDCKAPWRGPIFRVPIITVIKPIALS 768
423 G--PFIDSROQAGVNFNGTSHYERL-----IGAA--ELAQRASCYQ----- 459
769 GEPPALTLN-----LSFKSGHIERFINVPIGASWVEVTWRTSAFDTPRRFFLDTVOICP 824
460 ----KWQ-----VHRFARPEALGGTLHTNTIAGDLDA-----DFDI----- 490
825 LKRPKWEAVVTFSSPSLKNFSPFVEGGLTLELSIAQFWSSGIASHEPTCVDVEIIVPHGI 884
491 -----SLENDLKLKRVAEINAAQNPNNEVTVLLPQALQVGSPT 529
885 SVDQKIIGLDGSEAPVRVVARSLASERLVP-VATLNVKVTPTYPVRVESNLCSL-----PPS 939

Qy 530 HPSVPSGHATQNGAPATVVKALIGLDRGCEFP-----NPVFPSSDDGLE-----LINPEG 579
Db 940 RDRLPSPG---KQIIALTLYKFKLEDAEIKPRVPLNNRIY--DNKFESQYVIRSDSN 993
Qy 580 ACLTYEGEINKLVNVAFGROMIGHYRFDGQ 612
Db 994 KCVSYSGDVYPNYVKLSGEYTLQYIRHDNVQ 1026

RESULT 12

US-10-329-079-11
; Sequence 11, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; PRIOR FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 5245
; TYPE: PRT
; ORGANISM: Streptomyces fradiae
US-10-329-079-11

Query Match 3.3%; Score 116; DB 14; Length 5245;
Best Local Similarity 22.5%; Pred. No. 3.4;
Matches 159; Conservative 66; Mismatches 233; Indels 248; Gaps 41;
Qy 13 PMPTGVLR-LTSEORAKGWRROLEGEKSLGFHPSET---PYKYLEGSETKVKVLP 68
Db 3698 PRPE-----RULPSHAQRRLVFRQLEG-----PSATYVNVWALRUTG-----PL 3737
Qy 69 DGISASKILGIMARVRIATALAVLAAPCLAFDEVATAGVFPBEHKHTGEGHRLQCTN 128
Db 3738 DVAALRAALGDVVAR---HESLRVFA-----EDEL-----GPH-QVVL 3773
Qy 129 SDDALDPTAPNRRNVAF-----ASRRDAARRRGDTGVTCOITNGETDLATMFH 178
Db 3774 ADGPAPLSGPRVTDALPRLREADHAFRLDAEPPLR--AHLFATAPEHTLLVMHH 3831
Qy 179 KSLPHDELGO-----VTADDFAILE-----DCIL 202
Db 3832 --IATDWSORPLIADLAAYAAHAGRVFTLPPLPVAYADYALWQOARLGDREKDSAL 3889
Qy 203 NGDPSICEDVPAGDPAGRLVNPTAAFAIDISGPAFSATTIPPVPTLSSPELAAQLAELVW 262
Db 3890 SAQLAYWRDALAGSP-----EELALPADRPAPVPSHRGDSVPLTVPELHGRVVE--- 3940
Qy 263 MALARD---VPFMOY-----GTDEITTTAAAN-----LAGMGGFNLDVSI 301
Db 3941 --LAREFRATPFMVVHAALAAALLTRLGAGTDVPIGSPVAGRVDDALEDLVGF-FVNTLVL 3997
Qy 302 GSDCTVPP-PSOLF---RAIFVGV---ETGPF-----VSQLLVNSFT- 336
Db 3998 RTDTSGDPTFGELLERVRATDLGAYAHQDLPLFERLVELRDPERSLARHPLFQVSLNDYTA 4057
Qy 337 -----IDAITVEPK-----QETFPAD-----LNTMVD-FD 360
Db 4058 ETARDAARPELDGLTVSGRPLGVTTSKFDLTFALTETRAHDCGPGARGALEYSTDLFD 4117
Qy 361 EWLNIQGGPPAGPELDEELRFRINARDLARVFDVNIINTEAYRGSLLILELGAFRPG 420
Db 4118 R-----GTAERLAERPARVLQAAVAPGRLQIDV-LLPGRALLE-GEWSRPE 4165
Qy 421 INGFIFSDRQAGVNF-----GTSHY-----FRLIGAAELAQRASCYQKQVHRFAR 468
Db 4166 -PGVAPTD-DARPPDLFEAQARTPHAPAVRDGDRELSVAELNDRNRLARFLAARGAG 4223

Qy 469 PEALGGTLHNTIAGLDADDFDISLENDELKRVAEINA-----AQNPNNEVTVLL 519
Db 4224 PE-----DTVA-----VLLPRGPELITALTVAQKAGAAVYVPMDEALPABRIAHML 4268
Qy 520 ----PQAIQVSGSPHPSGSHATQNGAFATVVK-ALIGLDRGCEC 560
Db 4269 ENARPVLVLAHTATQDALPEGAGPVVRLDAPATEAALAGLD-GGDC 4313

RESULT 13

US-09-746-491-49
; Sequence 49, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USSN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-746-491-49

Query Match 3.2%; Score 114.5; DB 9; Length 912;
Best Local Similarity 19.8%; Pred. No. 0.24;
Matches 131; Conservative 82; Mismatches 232; Indels 215; Gaps 33;
Qy 107 SGVFPPEHKHTGGRHLQCTNSDDALDPTAPN-----RR 141
Db 278 NGTVPEGSK---PQTYVMVT-AIDADDPAQNAQMLRYILSQAPSSPSPNMFTINETG 333
Qy 142 DNVAFASSRRDAAR-----RERDGTGT-----VCQIT-----NGETDLATMFH 178
Db 334 DIITVAAGLDREKQVQYTLIIQATDMEGNPTYGLSNATAVITVTDNDNPPEETAMTFY 393
Qy 179 KSLPHDELQVTDADDFAILEDICILNGDFSCIEDVPAGDPAGR---LVNP-----TAAP 228
Db 394 GEVPENRVDVIVANLTVTDKDPHTPAWARYQMTGDPGTQGFILTDPNSNDGLVTVK 453
Qy 229 AID-----ISGPAFSATTIP-----PVPTLSSPEL 253
Db 454 PIDFETNRMFVLTVAAENQVPLAKGIQHPPOSTATVSIIVDVNESPYFVFPNPKLVROEE 513
Qy 254 AAQLAELYNMALARDVP-FMQYGTDEITTTAAANLAGMGGFPNLDVAVS--IGSDGTVDPP 310
Db 514 GLLAGSMLTFTTARDPPRYMQQ-----TSLRYSKLSPPANWLKIDPVNGQITTTAVLDRE 568
Qy 311 S-----QLFRATFVGVETG-PFVS-----QLLVNSFTIDAIVPEPKO-----ETAPDLNY 355
Db 569 SIYVQNNMYNATFLASDNGIPPMNSGTGLQIYLLIDINDNAPQVNPKEATTCTTLQPNAIN 628
Qy 356 MVDFDEWLNTQNGSP-----PAGPEELDEELRFRINARDLARVS----- 394
Db 629 ITAVDPDID-PNAGPFAFELPDPSPSIKRNWTVIRISGDHQAQLSLRIFLEAGYDVPIV 687
Qy 395 FVDNINTEAYRGSLLILELGAFRPGPFIDSDRQAGFVNFQGTSHYFR-LIGAAELAQ 453
Db 688 ITUSGNPHASSTSVLKVKV---CCQDINGDCTDVRIVG-AGLGTGAIILLCIILLI 743
Qy 454 RASCYQKQVHRFARPEAL-----CGLHNTIAGLDADDFDISLL 493
Db 744 LVLVFMVMMKRRDKERQAKLLIDPEDDVRDNLKLYDEEG-----GEEDQDYDLSQL 796
Qy 494 ENDELLKRVAEINAQNPNNEVTVLLQOATQ-VG-----SPTH-----SGHATQ 540
Db 797 QQPDTVB-----PDAIKPVIRLDERPIHAEPQYVRSAAHPHGD 837

```
QY 541 NGAFATVKALIGLDRGCEFPNVPFSDGGLLEINFEAGCLTYEGEINKLVNVAERQ 600
Db 838 IGDF--INEGLKAADN-----DPTAPPYDSLIVFDYEGSGST-AGSLSSINSSSGGEQ 888

RESULT 14
US-10-156-761-14890
; Sequence 14890, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14890
; LENGTH: 2365
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14890

Query Match 3.2%; Score 114; DB 14; Length 2365;
Best Local Similarity 22.6%; Pred. No. 1.4;
Matches 142; Conservative 56; Mismatches 233; Indels 198; Gaps 31;

QY 57 GSETWKKVLPDGTGISKILGKIMARVRIATALLVLAAPCLAFDEVTASGVFPPEHKK 116
Db 1162 GARTW-----IALVAESAELALRLR- AVAGEHAPLEGFTADDTTE 1202

QY 117 TGE-----GRHLOTCTNSDADLPTAPNRDNNVAFASRR 150
Db 1203 TGEVALLPFGGSGRQPCWAEFLFVAFPELQRLRLDETTARVLFP--PTAFDE--ASRK 1257

QY 151 DAARRRDGTGTCQITNGETDLATMFHKSILPHDELQVTTADDFAILDCILNGDFSICE 210
Db 1258 E--QOERITDTVAQPALGLTGLAAF--QLLTRAGIRPMAAGHSYGELAALAAAGALTTP 1313

QY 211 DVPAGDPAGRLVNPTAFAIDISGPAFSATTIPPVPTLSSPELAQAELAYMALARDVP 270
Db 1314 DALVTRSGR-----AAVIGAAGDGPDTM--AAVAGEPEVAALAAA----- 1356

QY 271 FMOYGTDEITTTAAANLAGMGFFNLDAVSGSDGTVDPSQLFRATFVGVTGP-----F 326
Db 1357 -----GLDGSVVTAEN-----SRQTVISGPTDDVLTAVERL-RAQGLGAKRIPVACAF 1405

QY 327 VSQLLVNSFTIDAITVEPKQTFAPDLNLM-----VDFDEWLNIQNGGPPAGPELDELRL 382
Db 1406 HSPL-----VAAAGEAFAKVLADVPVGATDFPVMNSNRTPAARYPESPEAIRAELA 1454

QY 383 FIRNARDLARVSFVDNINTEAYRSLILLELGAFS-----RPGINGPFIIDSRR 430
Db 1455 -----AQIGAPVRFPADQIEMYEAGARFVEAGPGSVLTRLVGTVGLDRPH-RTVALEKGR 1509

QY 431 QAGVNF-----GTSHYFRLLGAAELAQ-----RASCYQKMQV----- 463
Db 1510 HTGLVGLFLLAALQALAVAGVDVRTSWLFGQRCAGVDAARAPRVRAA-----VTVDGHLJRTA 1565

QY 464 -----HRFAR-PEALGGTGLNTIAGLDLADDFDLSLLENDELLKRVAEINAAQN--- 510
Db 1566 DGAIPAAALHPAERVPEAL---VTHSTSGAPVAATGSEALI--SEFLRTSREMIAAQRDVM 1621

QY 511 -----PNNEVTYLLPOAIOVGSPTHPSYPSGHAT-----QNGAF 544
Db 1622 LGYLGADEVPRPPAPAAAPATYADAPVTY-ADAPVTVTAP--PAADAAVAAPAGGSV 1678

QY 545 ATVLKALIGLDRGCEFPNVPFSDGGLLE 573
Db 1679 LSVLVDVIG-ERTG--YPVDMIEPDLDE 1704

RESULT 15
US-09-905-983-2
; Sequence 2, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Gallus gallus
US-09-905-983-2

Query Match 3.2%; Score 113.5; DB 9; Length 912;
Best Local Similarity 21.2%; Pred. No. 0.31;
Matches 113; Conservative 65; Mismatches 195; Indels 160; Gaps 28;

QY 175 TMFHKSILPH-----DELGOVTADDFAILEDCLNGDFESICEDVPAGPAGELVNP----- 224
Db 409 TVTDXDQOHPHTPAWNAHYQMTGGD-----PTGQFTILTDPNNSDGLVTVVXPIDPETN 460

QY 225 -----TAA-----FAIDISGPAFSATTIP-----PVPTLSSPELAQAEL 260
Db 461 RMFVLTVAAENQVPLAKGHOHPQSTATVSIITVIDVNESYFVFNPKLVQESGLLAGSM 520

QY 261 YMWALARDVP-FMOYGTDEITTTAAANLAGMGFFNLDAVS--IGSDGTVDPPS-----Q 312
Db 521 LTTFTARDPDRYMQQ-----TSLRYSKLSDPANWLKIDPVNGQITTTAVLDRESIYQNN 575

QY 313 LFRATFVGVTG-PFVS-----QLLVNSFTIDAITVEPKQ-----ETFAPDLNVMVDFDEW 362
Db 576 MYNATFLASDNGIPPMPSGTTGTLQIYLLDINDNAPQVNPKEATTCETLQPNAINITAVDPD 635

QY 363 LNIQNGGP-----PAGPEELDEELRFRNARDLARVS-----FVDNINT 401
Db 636 ID-PNAGPPAFAPFELDPSPSIKRWNTIVRISGDHQAQLSLRFRLEAGIYDVPVITDSGNP 694

QY 402 EAYRGSLLLELGAFAFRGPGINGPFIDSDROAGFVNFGTSHYFR-LIGAABELAQRASYQK 460
Db 695 HASSTSVLKVKV---CQCDDINGDCTVDRIVG-AGLGTGAIILLCIILLILVLMFVV 750

QY 461 QWHRFARPPAL-----CGTLHTIAGDLADDFDLSLLENDELLK 500
Db 751 WMKRRDKERQAKQLLIDPEDDVVRDNLKYDEEG-----GBEDQDYDLSUQQQPTVE 803

QY 501 RVAEINAAQNPNEVTYLLPOAIO-VG-----SPTH--PSYP-----SGHATQNGAFATV 547
Db 804 -----PDALKPVGIRLDERPIHAEPQYVRSAAHPGDIGDF--I 842

QY 548 LKALIGLDRGCEFPNVPFSDGGLLEINFEAGCLTYEGEINKLVNVAERQ 600
Db 843 NEGLKAADN-----DPTAPPYDSLIVFDYEGSGST-AGSLSSINSSSGGEQ 888

Search completed: September 19, 2004, 04:03:35
Job time : 137.581 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:52:28 ; Search time 29.3913 Seconds
(without alignments)
1187.397 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTRGSPMDTGV.....SIDGDMCGLVYTGADCOA 676

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3528	100.0	676	3	US-09-151-189-2
2	3528	100.0	676	4	US-09-596-794-2
3	129.5	3.7	3892	4	US-09-328-352-5503
4	123.5	3.5	2890	3	US-09-413-814-67
5	123.5	3.5	3798	3	US-09-335-409-6
6	123.5	3.5	3798	4	US-09-568-102-6
7	123.5	3.5	3798	4	US-09-567-969-6
8	123.5	3.5	3798	4	US-09-568-480-6
9	123.5	3.5	3798	4	US-09-568-486-6
10	123.5	3.5	3798	4	US-09-568-472-6
11	123.5	3.5	3798	4	US-09-567-899-6
12	122	3.5	22	3	US-09-151-189-9
13	122	3.5	22	4	US-09-596-794-9
14	117	3.3	1043	4	US-09-552-991A-32210
15	114	3.2	913	2	US-08-474-067-6
16	114	3.2	913	2	US-08-474-068A-6
17	114	3.2	913	2	US-08-472-481-5
18	113.5	3.2	3472	4	US-09-408-020-4
19	110	3.1	21	4	US-09-596-794-11
20	108.5	3.1	2315	4	US-09-543-681A-5434
21	108	3.1	19	3	US-09-151-189-10
22	108	3.1	19	4	US-09-596-794-10
23	102	2.9	440	4	US-09-684-855-101
24	102	2.9	440	4	US-09-684-855-124
25	102	2.9	440	4	US-09-684-855-146
26	102	2.9	466	3	US-08-868-435-2
27	102	2.9	466	3	US-08-868-435-12

28	102	2.9	466	4	US-08-744-231-2	Sequence 2, Appli
29	102	2.9	466	4	US-08-744-231-12	Sequence 12, Appli
30	102	2.9	466	4	US-09-636-499-7	Sequence 7, Appli
31	102	2.9	466	4	US-09-273-871A-12	Sequence 12, Appli
32	102	2.9	683	4	US-09-252-991A-33038	Sequence 33038, A
33	102	2.9	737	3	US-08-989-385-3	Sequence 3, Appli
34	102	2.9	737	4	US-09-593-826-3	Sequence 3, Appli
35	101.5	2.9	600	2	US-08-679-405-2	Sequence 2, Appli
36	101.5	2.9	600	2	US-08-842-799-2	Sequence 2, Appli
37	101.5	2.9	600	3	US-09-271-778-2	Sequence 2, Appli
38	101.5	2.9	600	4	US-09-788-871-2	Sequence 2, Appli
39	101.5	2.9	600	5	PCT-US96-11458-2	Sequence 2, Appli
40	101.5	2.9	654	4	US-09-543-681A-5710	Sequence 5710, Ap
41	100.5	2.8	564	4	US-09-252-991A-27329	Sequence 27329, A
42	100.5	2.8	1042	3	US-08-928-361B-11	Sequence 11, Appli
43	100.5	2.8	1042	4	US-09-588-995A-11	Sequence 11, Appli
44	100.5	2.8	1837	3	US-08-928-361B-5	Sequence 5, Appli
45	100.5	2.8	1837	4	US-09-588-995A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-151-189-2
; Sequence 2, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-09-151-189-2

Query Match						100.0%; Score 3528; DB 3; Length 676;
Best Local Similarity						100.0%; Pred. No. 0;
Matches						676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MLCHAADTRGSPMDTGVRLTLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET	60			
DB	1	MLCHAADTRGSPMDTGVRLTLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET	60			
QY	61	WKVKVLPDTGISASKILGKIMARVRIATALLVLAAPCLAFDEVTASGVPEEHKHTGEG	120			
DB	61	WKVKVLPDTGISASKILGKIMARVRIATALLVLAAPCLAFDEVTASGVPEEHKHTGEG	120			
QY	121	RHLQTCNDDALDPTAPNRDNVAFASRRDARRRDGTGTVCQITNGETDLATMPHKS	180			
DB	121	RHLQTCNDDALDPTAPNRDNVAFASRRDARRRDGTGTVCQITNGETDLATMPHKS	180			
QY	181	LPHELGQVTDADPAILEDICILNGDPSICEDVPAGDPAGRLVNPPTAAFAIDISGPAFSAT	240			
DB	181	LPHELGQVTDADPAILEDICILNGDPSICEDVPAGDPAGRLVNPPTAAFAIDISGPAFSAT	240			
QY	241	TIPPVPTLSPPELAOLAELYNMALARDVPFMOYGYTDEITTTAAANLAGMGFPNLDVDS	300			
DB	241	TIPPVPTLSPPELAOLAELYNMALARDVPFMOYGYTDEITTTAAANLAGMGFPNLDVDS	300			
QY	301	IGSDGTVDFPSQLFRATFVGVEVGTFVSQLVNSFTIDAITVPEPKQETFAFDLNMVDFD	360			
DB	301	IGSDGTVDFPSQLFRATFVGVEVGTFVSQLVNSFTIDAITVPEPKQETFAFDLNMVDFD	360			
QY	361	EWLNQNGGPPAGPEELDELRIRNARDLARVSVVDNINTEAYRGLSLILLELGAFSRPG	420			

Db 361 EWLNIQNGPPAGPEELDEBLRIFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420
Qy 421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGLHNTI 480
Db 421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGLHNTI 480
Qy 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSYPSGHATQ 540
Db 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSYPSGHATQ 540
Qy 541 NGAFATVVKALIGLDRGCECFPNPVPFSDGGLLELINFEGACLYEGBEINKLAVNVAFRQ 600
Db 541 NGAFATVVKALIGLDRGCECFPNPVPFSDGGLLELINFEGACLYEGBEINKLAVNVAFRQ 600
Qy 601 MLGIHYRFDDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660
Db 601 MLGIHYRFDDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660
Qy 661 DMCSGLVYTGADCOA 676
Db 661 DMCSGLVYTGADCOA 676

RESULT 2

US-09-596-794-2
; Sequence 2, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-09-596-794-2

Query Match 100.0%; Score 3528; DB 4; Length 676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLCHAADTTGSPMPDTGVRLTLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET 60
Db 1 MLCHAADTTGSPMPDTGVRLTLTSEORAKGWRQLEGEKSLGFHPSETPYIKYLEGSET 60
Qy 61 WKVKLPDTGISASKILGKIMARVRIATALAVLAAAPCLAFDEVTAGVFPPEEHKHTGEG 120
Db 61 WKVKLPDTGISASKILGKIMARVRIATALAVLAAAPCLAFDEVTAGVFPPEEHKHTGEG 120
Qy 121 RHLQCTNSDDALDPTAPNRDNNVAFASRRDAARRRDGTGTCQITNGETDLATMPHKS 180
Db 121 RHLQCTNSDDALDPTAPNRDNNVAFASRRDAARRRDGTGTCQITNGETDLATMPHKS 180
Qy 181 LPHDELQVTDADDPFAILEDCLINGDPSICEDVPAGDPAGRLVNPTAAFAIDISGAPFSAT 240
Db 181 LPHDELQVTDADDPFAILEDCLINGDPSICEDVPAGDPAGRLVNPTAAFAIDISGAPFSAT 240
Qy 241 TTPVPVTLSSPELAAQLAELYNALARDVFPFQYGTDEITTTAAANLAGMGFPNLDVAVS 300
Db 241 TTPVPVTLSSPELAAQLAELYNALARDVFPFQYGTDEITTTAAANLAGMGFPNLDVAVS 300
Qy 301 IGSDGTVDPSQLFRATFVGVETGPFVSQLLVNSFTTIDAITVEPKQETFPADLNVMVDFD 360
Db 301 IGSDGTVDPSQLFRATFVGVETGPFVSQLLVNSFTTIDAITVEPKQETFPADLNVMVDFD 360

Qy 361 EWLNIQNGPPAGPEELDEBLRIFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420
Db 361 EWLNIQNGPPAGPEELDEBLRIFIRNARDLARVSFVDNINTEAYRGSLLILLELGAFSRPG 420
Qy 421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGLHNTI 480
Db 421 INGPFFIDSDRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRFARPEALGGLHNTI 480
Qy 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSYPSGHATQ 540
Db 481 AGDLDDADFDISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSYPSGHATQ 540
Qy 541 NGAFATVVKALIGLDRGCECFPNPVPFSDGGLLELINFEGACLYEGBEINKLAVNVAFRQ 600
Db 541 NGAFATVVKALIGLDRGCECFPNPVPFSDGGLLELINFEGACLYEGBEINKLAVNVAFRQ 600
Qy 601 MLGIHYRFDDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660
Db 601 MLGIHYRFDDGIQGLLGETITVTRLHOELMTFAEATFEPRFLTGEVVKLFQDGTFSIDG 660
Qy 661 DMCSGLVYTGADCOA 676
Db 661 DMCSGLVYTGADCOA 676

RESULT 3

US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
; LENGTH: 3892
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5503

Query Match 3.7%; Score 129.5; DB 4; Length 3892;
Best Local Similarity 21.0%; Pred. No. 0.0053;
Matches 158; Conservative 71; Mismatches 245; Indels 279; Gaps 36;
Qy 46 PSETPYIKYLEGSE-----TWKKVLPDGTGISASKILGKIMARVRIAT----- 88
Db 330 PGSTVTVTVYDGTGSKTVVAGPDGTW---TVNPNGLNDGDEV-----TAVATDPAGNTSG 380
Qy 89 -ALAVVLA-APCLAFDEV-----TASGV-----FPEEHKHTG----- 118
Db 381 PATAVVDVAAPTVALDVLNDSTPALTGTVNDPTATVTVNVDGVDYPAVNNGDGTWTLA 440
Qy 119 -----EGRHLQTCNSDDA-----LDPTAPN-----RDNVAFAS 148
Db 441 DNTLPTLADGPHITITATDAAGNVGTDGTVTVTAAPNTAGVTTFIDSVTADNVAS 500
Qy 149 RRDAARR-----ERDGTGT-VCOITNGETDLATM-----FHKS LPHDELQVTA 191
Db 501 --EAGNVNITIGVKLNIPADATNTAVTVINGVTYNATVDKTAGTWTVSVEGSL----- 553
Qy 192 DDFAILDCILNGDPSICEDVPAGDPAGR--LVNPTAAFAIDISGAPFSAITTPVPVTL 249
Db 554 -----VADAKTIDAKVTFTDAAGNSSTVNDTQIYTLDTAAAP--APVIDPV----- 598
Qy 250 SPELAAQLAELYNALARDVFPFQYGTDEITTTAAANLAGMGFPNLDVAVS1--GSDGTV 307
Db 599 -----NGTDPITGTAEFGSTVTVTVYPNGDTATVAGPDGWSW 634
Qy 308 D-PFSQLFRATFV-GVETGPFVSQLLVNSFTTIDAITVEPKQETFPADLNVMVDFEWLNI 365

Db 635 SVNPGLNDGDEVEIAITDPAGNPSLPGTATVDA--VGPNTD-----GVNFTVDSVTADNV 688
Qy 366 QNGGPPAGPELDELRFIRNARDLARVSVDNINTEAYRGSLLILBELGAFSRPGINGPF 425
Db 689 INASEASGNVTVTGLXNV--PADAANTVTVVINGQTYTATV----- 729
Qy 426 IDSRQAGFVNGTSHYFRLLIGAAELAQASCYQKQVHRPARPEALGGTLHNTIAGDLD 485
Db 730 ---DSTAG-----TWTV-----SVPGSDLT 746
Qy 486 ADPDISLENDELKRVAEINAAONP---NNEVTYLL---PQA-----I 523
Db 747 ADAD-----KTIDAKVTFTDAAGNSSVNDTQYITTTAPDAPVNPVNGTDPTGTA 800
Qy 524 QVGSPTSPSYSGHATQGAFAATVLKALIGLDRGCEFPNPFPSDDGLBLINPEGACLT 583
Db 801 EPGSTVTVTPDGSTT-----TVVAGPD-GTWTVPNPLNDG----- 837
Qy 584 YEGEINKLVANVAPGRQMLGIHYRFDGIGQLLIGETTIVTTLHQLMTFAEATFERPLP 643
Db 838 ---KVTIAITDPA--GNPSLPGTATVDAVGPNTDGVNFTVDSVTADNVINASEASGNVTV- 892
Qy 644 TGEVILFQDGTFSIDGDMCSGLVYTVGVADCOA 676
Db 893 TGLXNVPPADAANTVTVVINGQTYTATVDSA 925

RESULT 4

US-09-413-814-67
; Sequence 67, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Bayer, Stefan
; APPLICANT: Blocker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 95/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-67

Query Match 3.5%; Score 123.5; DB 3; Length 2890;
Best Local Similarity 21.9%; Pred. No. 0.014;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;
Qy 158 DGTGTCQITNGETDLMTHFKSLPHDELQGVTDADFAILEDG----- 200
Db 1989 DGEHFVSSQPLPEPDLAALV-----EAGRVFADLPVLEWCKFAGERLADVLTKTLA 2042
Qy 201 --1L--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2043 LEILFPGSGFDMAERIYRDSPIARYSNGIVGVVESAARVAVPSGMSILEIGAGTGATT 2102
Qy 242 IPPVPTLSSPELAQAELAYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295
Db 2103 AAVLPVLLPDRTEYHFTDVSPFLARAEQRFRDYPFLKYGLDQDPAGQYAHQRFDV 2162

Qy 296 LDAVSGSDGTVPFSQLFRATFVGTVGTFVPSQLLVNSFTIDAITVEPKQETPAP-DLN 354
Db 2163 IVAANV-IHATRD-----IRAT-----AKRL-----SLAPGGLL 2192
Qy 355 YMYD---FDEWLNQNGPPAGPEELDEELRF-----IRNARD-----LARVSFVDNI----- 399
Db 2193 VLVEGTGHPITWFIDTT-GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 2251
Qy 400 -NTEAVRGSLLILELGAFFSRPGINGPPIIDSDROAGFVNGTSHYFRLLIGAAELAQASCY 458
Db 2252 GSPAGILQGVHLS-----RAGIAGAACDSGES-----ATESPAARAVR 2292
Qy 459 QKWQ-----VHRFA-----RPEALGGLTHTNTIAGDL-----DADFDI 490
Db 2293 QEWADGSADVHRMALERMYFHRPRGRQVWVHGRLRTGGGAFTKALAGDLLLFFDTGQV 2352
Qy 491 SILENDELKRVAEINAAQNNEVTYLLPQAQVGSPTSPSYSGHAT-----QNG 542
Db 2353 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDP:PEAPAAASSSSAGAWLVLMDOGG 2412
Qy 543 AFATVLKALIGLDRGGEC 560
Db 2413 TGAALVSLLEG--RGENC 2428

RESULT 5

US-09-335-409-6
; Sequence 6, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-6

Query Match 3.5%; Score 123.5; DB 3; Length 3798;
Best Local Similarity 21.9%; Pred. No. 0.022;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;
Qy 158 DGTGTCQITNGETDLMTHFKSLPHDELQGVTDADFAILEDG----- 200
Db 2760 DGEHFVSSQPLPEPDLAALV-----EAGRVFADLPVLEWCKFAGERLADVLTKTLA 2813
Qy 201 --1L--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILFPGSGFDMAERIYRDSPIARYSNGIVGVVESAARVAVPSGMSILEIGAGTGATT 2873
Qy 242 IPPVPTLSSPELAQAELAYMALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295
Db 2874 AAVLPVLLPDRTEYHFTDVSPFLARAEQRFRDYPFLKYGLDQDPAGQYAHQRFDV 2933
Qy 296 LDAVSGSDGTVPFSQLFRATFVGTVGTFVPSQLLVNSFTIDAITVEPKQETPAP-DLN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLAPGGLL 2963
Qy 355 YMYD---FDEWLNQNGPPAGPEELDEELRF-----IRNARD-----LARVSFVDNI----- 399
Db 2964 VLVEGTGHPITWFIDTT-GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022

Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 8

US-09-568-480-6

; Sequence 6, Application US/09568480

; Patent No. 6355458

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,480

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 3798

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-480-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;

Best Local Similarity 21.9%; Pred. No. 0.022;

Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

Qy 158 DGTGTCQITNGETDLATMFHKSPLHDELQGVTDADDFAILEDG----- 200

Db 2760 DGEHFVSSQPLPEPDLAALV-----EAGRVFADLPVLFWECKFAGERLADVLTKTLA 2813

Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGAPFATT 241

Db 2814 LEILPFGSGFDMAERIVRDSPIARYSNGIVRGVVEAARVAVPAGMFSILEIGAGTGATT 2873

Qy 242 IPPVPTLSSPELAQALAEYMWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295

Db 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEQRFDRYPFLKYGLLDVDDQEPAGQYAHQRFDV 2933

Qy 296 LDVAVSIGSDGTVDPFSQLFRATFVGTGPFVSGVQLLVNSETIDAITVEPKQETPAP-DLN 354

Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963

Qy 355 YMYD---FDEWLNIQGGPPAGPEELDEELRF---IRNARD---LARVSFVNI----- 399

Db 2964 VLVEGTGHPITWFDITT-GLIEGMQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022

Qy 400 -NTEAYRGSLLLELGAFSRPGINGPFIQSDROAGFVNFCTSHYFRILIGAAELAQASCY 458

Db 3023 GSPAGILGQHVILS-----RAGIAGAACDSSGES-----ATESPAARAVR 3063

Qy 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490

Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGRLRTGGGAFTKALAGDLLLFDGTQGVV 3123

Qy 491 SLENDELLKRVAEINAAQNPNNEVYLLPQATQVGSPTSPSPSGHAT-----QNG 542

Db 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDPIPEAPAAASSSSAGAWLVLMQDGG 3183

Qy 543 AFATVLKALIGLRGGEC 560

Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 9

US-09-568-486-6

; Sequence 6, Application US/09568486

; Patent No. 6355459

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

; FILE REFERENCE: 4-30582A

; CURRENT APPLICATION NUMBER: US/09/568,486

; CURRENT FILING DATE: 2000-05-10

; PRIOR APPLICATION NUMBER: 09/335,409

; PRIOR FILING DATE: 1999-06-17

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 3798

; TYPE: PRT

; ORGANISM: Sorangium cellulosum

US-09-568-486-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;

Best Local Similarity 21.9%; Pred. No. 0.022;

Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

Qy 158 DGTGTCQITNGETDLATMFHKSPLHDELQGVTDADDFAILEDG----- 200

Db 2760 DGEHFVSSQPLPEPDLAALV-----EAGRVFADLPVLFWECKFAGERLADVLTKTLA 2813

Qy 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAAFAIDISGAPFATT 241

Db 2814 LEILPFGSGFDMAERIVRDSPIARYSNGIVRGVVEAARVAVPAGMFSILEIGAGTGATT 2873

Qy 242 IPPVPTLSSPELAQALAEYMWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFPN 295

Db 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEQRFDRYPFLKYGLLDVDDQEPAGQYAHQRFDV 2933

Qy 296 LDVAVSIGSDGTVDPFSQLFRATFVGTGPFVSGVQLLVNSETIDAITVEPKQETPAP-DLN 354

Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963

Qy 355 YMYD---FDEWLNIQGGPPAGPEELDEELRF---IRNARD---LARVSFVNI----- 399

Db 2964 VLVEGTGHPITWFDITT-GLIEGMQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022

Qy 400 -NTEAYRGSLLLELGAFSRPGINGPFIQSDROAGFVNFCTSHYFRILIGAAELAQASCY 458

Db 3023 GSPAGILGQHVILS-----RAGIAGAACDSSGES-----ATESPAARAVR 3063

Qy 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490

Db 3064 QEWADGSADVHRWALERMVYFHRPPGQVWVHGRLRTGGGAFTKALAGDLLLFDGTQGVV 3123

Qy 491 SLENDELLKRVAEINAAQNPNNEVYLLPQATQVGSPTSPSPSGHAT-----QNG 542

Db 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDPIPEAPAAASSSSAGAWLVLMQDGG 3183

Qy 543 AFATVLKALIGLRGGEC 560

Db 3184 TGAALVSLLEG--RGEAC 3199

RESULT 10

US-09-568-472-6

; Sequence 6, Application US/09568472

; Patent No. 6358719

; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas

; APPLICANT: Ligon, James

; APPLICANT: Molnar, Istvan

; APPLICANT: Zirkle, Ross

; APPLICANT: Cyr, Devon

; APPLICANT: Goerlach, Joern

; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;
Best Local Similarity 21.9%; Pred. No. 0.022;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;
QY 158 DGTGTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG----- 200
DB 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFEWCKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAFAIDISGPAFSATT 241
DB 2814 LEILPFGGSFDMARIYRDSPIARYSNGIVRGVVEAARVAVPSGMFSILEIGAGTGATT 2873
QY 242 IPPVPTLSSPELAQAELAYMWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295
DB 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEORFRDYPFLKYGILDVDQEPAGQGYAHORFDV 2933
QY 296 LDVAVSIGSDGTVPFSPQLFRATFVGTVGTGPFVSQLVNSFTIDAIVPEPKQETFAP-DLN 354
DB 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNIQNGPPAGPEELDELRP---IRNARD-----LARVSFVDNI----- 399
DB 2964 VLVEGTGHPITWFDITT-GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022
QY 400 -NTEAYRGSILILELGAFAFRPGINGPFIDSDROAGFVNFOTSHYFRILGAELAQRASY 458
DB 3023 GSPAGILGQHVILS-----RAGIAGACDSGSES-----ATESPAARAVR 3063
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
DB 3064 QEWADGSADVHRMALERMYFHRPGRQVVMVHGRLRTGGAFYKALAGDLLLFDITGQVV 3123
QY 491 SLENDELLKRVAEINAAQNPNNEVYLLPQAIQVGSPTSPSPSGHAT-----QNG 542
DB 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDIPEAPAAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
DB 3184 TGAALVSLLEG--RGEAC 3199

RESULT 11
US-09-567-899-6
; Sequence 6, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-6

Query Match 3.5%; Score 123.5; DB 4; Length 3798;
Best Local Similarity 21.9%; Pred. No. 0.022;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;
QY 158 DGTGTCQITNGETDLATMFHKSPLPHDELQGVTTADDFAILEDG----- 200
DB 2760 DGEHFVSSQPLPEPDLAVAL-----EAGRVFADLPVLFEWCKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDFSICEDVPAGDP-----AGRLVNPTAFAIDISGPAFSATT 241
DB 2814 LEILPFGGSFDMARIYRDSPIARYSNGIVRGVVEAARVAVPSGMFSILEIGAGTGATT 2873
QY 242 IPPVPTLSSPELAQAELAYMWALA-----RDVPFMOYGTDEITTTAAANLAGMGGFNP 295
DB 2874 AAVLPVLLPDRTEYHFTDVSPLFLARAEORFRDYPFLKYGILDVDQEPAGQGYAHORFDV 2933
QY 296 LDVAVSIGSDGTVPFSPQLFRATFVGTVGTGPFVSQLVNSFTIDAIVPEPKQETFAP-DLN 354
DB 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNIQNGPPAGPEELDELRP---IRNARD-----LARVSFVDNI----- 399
DB 2964 VLVEGTGHPITWFDITT-GLIEGWQKYEDDLRIDHPLLPARTWCDVLRVGFADAVSLPGD 3022
QY 400 -NTEAYRGSILILELGAFAFRPGINGPFIDSDROAGFVNFOTSHYFRILGAELAQRASY 458
DB 3023 GSPAGILGQHVILS-----RAGIAGACDSGSES-----ATESPAARAVR 3063
QY 459 QKWQ-----VHRFA-----RPEALGGTLHNTIAGDL-----DADFDI 490
DB 3064 QEWADGSADVHRMALERMYFHRPGRQVVMVHGRLRTGGAFYKALAGDLLLFDITGQVV 3123
QY 491 SLENDELLKRVAEINAAQNPNNEVYLLPQAIQVGSPTSPSPSGHAT-----QNG 542
DB 3124 AEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDIPEAPAAAASSSSAGAWLVLMDOGG 3183
QY 543 AFATVLKALIGLDRGGEC 560
DB 3184 TGAALVSLLEG--RGEAC 3199

RESULT 12
US-09-151-189-9
; Sequence 9, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473
US-09-151-189-9
Query Match 3.5%; Score 122; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 AQRASYQKQVHRFARPEALG 473
|||||
Db 1 AQRASYQKQVHRFARPEALG 22

RESULT 13

US-09-596-794-9
; Sequence 9, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,
; OTHER INFORMATION: 1st conserved motif
US-09-596-794-9

Query Match 3.5%; Score 122; DB 4; Length 22;

Best Local Similarity 100.0%; Pred. No. 3.2e-06;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 AQRASYQKQVHRFARPEALG 473
|||||
Db 1 AQRASYQKQVHRFARPEALG 22

RESULT 14

US-09-252-991A-32210
; Sequence 32210, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32210
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32210

Query Match 3.3%; Score 117; DB 4; Length 1043;

Best Local Similarity 20.1%; Pred. No. 0.011;

Matches 138; Conservative 83; Mismatches 257; Indels 208; Gaps 34;

QY 52 IKYLEG-----SETWKKVLPDGTGASAKILGKIMAR-----VRIATATAVVL 94
|||||
Db 460 VDLDGDCDPQIVTGRTRATNRPYALPDHKILUSTKSKYKGSRAEIRIDDTTAAIS 519
|||||

QY 95 AAPCLAFDEVTASGVF-----PEEHKHTGEGRLTCTNSDDALDPTAPNRDNVAF 146

Db 520 AA---LMSDRGASALHLGYLTHPRPEGKPRGEGFELRT-----DEHGAVRAAKGLL 568
QY 147 AS-----RRDAARRRDTGTVCQITNGETDLATMF-----HKSPLPHDLGQVTTADDDFA 195
Db 569 LSTEEQLRAGAGHLDLDR---GVVVQVLEAALEARELDGYAGEHQGVGHDAAPQOT----- 620
QY 196 ILEDICILNGDFSICEDV--PAGDPAGRLVNPTAAFAIDISGPAPSATIIPVPTLSSPEL 253
Db 621 -LQEA-----RDLGHGANDSGSKSGKGAIPA--LSGPAGIAAATPASITLAAAGEH 669
QY 254 AAQLAELYWALARDVPMQYGTDEITTTAAANLAGMGFPNLDVAVIGSDGTVDPSQL 313
Db 670 VDSVARQNOQVTAQKVVINAGSD-----IGLFAQGG-----EL 703
QY 314 FRATFVGVEGPFVSOLLVNSFTIDAITVEPKOBTAPDLNMYVDFEWLNIQNGGPPAG 373
Db 704 RQIT-----HOGPMLLQAKNDIRLEA---EQSEVSASQOHVLTAKHITLMCGG----- 752
QY 374 PEELDEELRFIRNARDLARVSFVDNINTEAY---RGSLLLELGAFAFRPGINGPFIDSDR 430
Db 753 -----AYLTLLKGGNI--ELG---MPG----- 768
QY 431 QAGFVNFGTSHYFRLIGAELAORASCYQKQVHRFARPEALG-----TLHN- 478
Db 769 --NFVVKAARKSH--VGAASLEAELEPQFVEGTQRREVLKQLDGTAMPNVPYITTMANG 824
QY 479 -TTAGDLDDADFDISLENDLLKRVAEINAAQNPNNVTVL---LPOAIOVGS--PTHPSY 533
Db 825 EVIEGVTDAEGATQLQKDMNTAKVDMKTKSPASAVAGIAAAGAAVAVGKLLSGFDA 884
QY 534 PSGHATQNGAFATVLKALIGLDRGECFPNPFPSDDGLELINF--EGACLTYEIGEINKLA 592
Db 885 EAGRALSEGE-----ISLANG-----VFGDSIDYSTVRLRDRDYVPWQCKDYMA 929
QY 593 VN--VAFGRQMLGI-HYRFDGI--QGLLL-----GEITVTLHQLMTFAEEA 636
Db 930 PNGHIYFGEELRGVADWSLESQRLQGLFIHEMTHVMOHQGVNVLVGVAYQQAQRFLLGD 989
QY 637 TFEFRFLTGEVVKLFQDGTFSIDGDM 662
Db 990 QYAYRLEPGTKL---DYNIEQQGDI 1012

RESULT 15

US-08-474-067-6
; Sequence 6, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/474,067
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293

```
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-474-067-6

Query Match      3.2%; Score 114; DB 2; Length 913;
Best Local Similarity 21.2%; Pred. No. 0.018;
Matches 113; Conservative 65; Mismatches 196; Indels 159; Gaps 28;

Qy 175 TMFHKSLPH---DELGQVTADDFAILEDCLINGDFSCIEDVPAGDPAGRLVNP----- 224
Db 409 TVTDKQDPHTFAMWARYQMTGGD-----PTGQFTILTDPNNDGLVTVVKPIDFTN 460
Qy 225 ----TAA-----FAIDISGPAFSATTIP-----PVPTLSSPELAAQLAEL 260
Db 461 RMFVLTVAAENQVPLAKGIQHPPQSTATVITVIDVNESPVFVFNPKLVQEGLLAGSM 520
Qy 261 YWMALARDVP-FMQYGTDEITTTAAANLAGMGFPNLDVAVS--IGSDGTVDPPS-----Q 312
Db 521 LTTFTARDPORYMQQ-----TSLRYKSLSPANWLKIDPPVNGQITTTAVLDRESIYVQNN 575
Qy 313 LFRATFVGVEVG-PFVS-----QLLVNSFTIDAITVEPKQ-----ETFAPDLNVMVDFDEW 362
Db 576 MYNATFLASONGIPMSGTGTLYLLDINDNAPQVNPKEATTCETLQPNAINITAVDDPD 635
Qy 363 LNIQNGGP-----PAGPEELDEELRIFIRNARDLARVS-----FVDNINT 401
Db 636 ID-PNAGPFAPFELPDSPPSIKRNWTVIRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNP 694
Qy 402 EAYRGSLLLELGFAPRPGINGPIDSDROAGFVNFQTSHYFR-LIGAAELAQKASCYQK 460
Db 695 HASSTSVLKVKV---CQCDINGDCTDVRIVG-AGLGTGAI IAILLCIIILLILVLMFVV 750
Qy 461 WQVHRFARPEAL-----GGTLHTIAGDLADDFDISLLENDELLK 500
Db 751 WMKRDKERQAKQLLIDPDVDDVRNLIKYEDEGG-----GEEDQYDLSQLQOPTVE 803
Qy 501 RVABEINAAQNPNEVYLLPQAIQ-VG-----SPTH--PSYP-----SGHATONGAFATV 547
Db 804 -----PDAIKPVGIRRLDERPIHAEPQYPVRSAAHPHGDIGDFINE 844
Qy 548 LKALIGLDRGCEFPNPVFFSDGDLINFEACLYTEGEINKLVANVAFGRQ 600
Db 845 ----GLAKAAD--NDPTAPPYDSLLVFDYEGSGST-AGSLSSLNSSSSGGEQ 889
```

Search completed: September 19, 2004, 04:00:32
Job time : 31.3913 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 04:03:44 ; Search time 1014 Seconds
(without alignments)
3364.964 Million cell updates/sec

Title: US-10-691-383-2
Perfect score: 3528
Sequence: 1 MLCXAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVAQCQA 676

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10691383/runat.17092004.102657.1795/app_query.fasta_1.1230
-DB=Published Applications NA -QFMT=fastap -SURFIX=rnnpb -MINMATCH=0.1
-LOOPCUT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-MAXLEN=2000000000 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
19: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

RESULT 1
US-09-840-762A-1
; Sequence 1, Appli US-09-840-762A-1
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840, 762A.
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 09/151,189
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)

ALIGNMENTS

1	3528	100.0	2931	9	US-09-840-762A-1	Sequence 1, Appli
2	3528	100.0	2931	17	US-10-691-383-1	Sequence 1, Appli
3	131	3.7	9579	15	US-10-132-134-11	Sequence 11, Appli
4	131	3.7	52101	15	US-10-132-134-1	Sequence 11, Appli
5	129	3.7	1236	17	US-10-437-963-29553	Sequence 29553, A
6	128.5	3.6	1953	16	US-10-369-493-42356	Sequence 42356, A
7	124.5	3.5	7185	15	US-10-329-079-48	Sequence 34, Appli
8	124.5	3.5	61944	13	US-10-329-079-34	Sequence 34, Appli
9	124	3.5	18930	13	US-10-282-122A-31609	Sequence 31609, A
10	123.5	3.5	68750	14	US-10-014-717-1	Sequence 1, Appli
11	123	3.5	5820	13	US-10-152-886-44	Sequence 44, Appli
12	122	3.5	2843	13	US-10-194-163-21	Sequence 21, Appli
13	122	3.5	3969	15	US-10-156-761-2515	Sequence 2515, Ap
14	122	3.5	9025608	15	US-10-156-761-1	Sequence 1, Appli
15	121	3.4	5040	17	US-10-437-963-65478	Sequence 65478, A
16	120	3.4	2664	15	US-10-156-761-1241	Sequence 1241, Ap
17	118	3.3	2256846	17	US-10-470-585-1	Sequence 1, Appli
18	117.5	3.3	3505	13	US-10-282-122A-23847	Sequence 23847, A
19	117	3.3	53500	10	US-09-953-348-76	Sequence 76, Appli
20	117	3.3	53500	15	US-10-267-255-76	Sequence 1, Appli
21	116.5	3.3	1830121	15	US-10-329-960-1	Sequence 1, Appli
22	116.5	3.3	1830121	16	US-10-329-670-1	Sequence 1, Appli
23	116.5	3.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
24	116	3.3	5352	17	US-10-437-963-29803	Sequence 29803, A
25	116	3.3	15738	15	US-10-329-079-12	Sequence 12, Appli
26	116	3.3	37360	15	US-10-329-079-6	Sequence 6, Appli
27	115	3.3	2193	17	US-10-437-963-19583	Sequence 19583, A
28	115	3.3	64492	13	US-10-378-083-1	Sequence 1, Appli
29	115	3.3	125401	17	US-10-203-293-35	Sequence 35, Appli
30	114	3.2	972	17	US-10-317-271A-12	Sequence 12, Appli
31	114	3.2	7095	15	US-10-156-761-7340	Sequence 7340, Ap
32	114	3.2	11007	16	US-10-402-842-5	Sequence 5, Appli
33	114	3.2	11402	17	US-10-672-396-10	Sequence 10, Appli
34	114	3.2	47988	16	US-10-402-842-1	Sequence 1, Appli
35	114	3.2	2256846	17	US-10-470-585-1	Sequence 1, Appli
36	113.5	3.2	2586	16	US-10-369-493-31958	Sequence 31958, A
37	113.5	3.2	2824	9	US-09-905-983-1	Sequence 1, Appli
38	113.5	3.2	2824	9	US-09-905-983-3	Sequence 3, Appli
39	113.5	3.2	10419	14	US-10-027-806-3	Sequence 3, Appli
40	113.5	3.2	10419	14	US-10-034-623-3	Sequence 3, Appli
41	113.5	3.2	10419	15	US-10-027-801-3	Sequence 3, Appli
42	113.5	3.2	10419	15	US-10-029-120-3	Sequence 3, Appli
43	113.5	3.2	42432	14	US-10-027-806-2	Sequence 2, Appli
44	113.5	3.2	42432	14	US-10-034-623-2	Sequence 2, Appli
45	113.5	3.2	42432	15	US-10-027-801-2	Sequence 2, Appli

OTHER INFORMATION: vanadium bromoperoxidase
US-09-840-762A-1

Alignment Scores:

Pred. No.: 0 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-691-383-2 (1-676) x US-09-840-762A-1 (1-2931)

```
QY 1 MetLeuCysHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20
DB 228 ATGCTTTGGCATGCGAGCGGACACGACGAGAGGCTCTCTATGCTGTGACACCGAGTGCTT 287
QY 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyLys 40
DB 288 CGGTTGCTCATCATCAGACGAGCGGCTAAGGTTGGAGACGCCAGTTAGAGGGGGAGAA 347
QY 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
DB 348 TCACCTAGGTTTTCATCAAGCGAGACGCTTATATCAAGTACTTGGAAAGGCTCTGAGACT 407
QY 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
DB 408 TGAAGAAGGTTAAGCTTCCAAAGCGAGCGGATATCGGCTTCCAAAGATCCCTGGGTAAAAAT 467
QY 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
DB 468 ATGGCCAGGTCGCGATCGCTACCGCTTGGCTGTGCTACTGGCCGACACCTGTTGGCA 527
QY 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
DB 528 TTCGACGAGGTCACAGCAGTGGTGTGTTCCCTCAGGAAACACAAAGCACACCGCGGGGGA 587
QY 121 ArgHisLeuGlnThrCysThrAnSerAspAlaLeuAspProThrAlaProAsnArg 140
DB 588 AGACACCTCCAGACCTGTGACAACTCCGACGATGCGCTGGATCCGACGCGCGCAATAGA 647
QY 141 ArgAspAsnValAlaPheAlaSerArgArgAlaAlaAlaArgArgGluArgAspGlyThr 160
DB 648 AGGACAACTAGCTTTTGGTCGCGCGCGGATGCCGCCAGGCGAGAACGCTGACGGGACA 707
QY 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
DB 708 GGGACTGCTCTGCCAAATCACTAAGCGAGAACTGATTTGGCTTACCATGTTCCCAAGTCT 767
QY 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
DB 768 CTGCCACACGATGAATCGGACAGGTAAACCGACAGACATTCGCTATCTCTGAGGACTGC 827
QY 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
DB 828 ATCTTAAACGGAGATTTTCAGCAATTTGCGAGACGTGCTCGCGGAGACCGCGGGGTGCG 887
QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
DB 888 CTGCTCAATCTTACCGCTGGGCTTGGCCATCGACATATCCGGTCCCGCATTTCTCGGCTACG 947
QY 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
DB 948 ACAATACCCCGGTACCTACCTTTCCTCTGAGCTCCCGCTCAGTTGGGGAGGCTA 1007
QY 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
DB 1008 TACTGGATGGGCTGGCCAGGATGTACCTTTATGCAATATGGCACCGAGAAATTACC 1067
QY 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
DB 1068 ACTACCGCGGACCAACCTCGCTGGAATGGGAGGCTTCCAAATCTGGACGCGCTGTGCG 1127
```

```
QY 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
DB 1128 ATAGGGTCGATGGTACGGTGGACCGGTTCTCCAGCTCTTCGAGGACCTTCGTTGGT 1187
QY 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
DB 1188 GTTGAACCGGGGCCCTTTGTCTCTAGCTGCTCGTGAACAGCTTACCATCGACGCTATT 1247
QY 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
DB 1248 ACGGTGGAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGGTCGATTTGAC 1307
QY 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
DB 1308 GAATGCTGAACATTCAGATGTTGGACCCCGCGCGGCCCGAGAGATTAGACGAAGAG 1367
QY 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
DB 1368 CTGCGTTTATTCGTAACGCGCGACCTGGCCGAGGCTCTCTTCGTGGACAATATCAAC 1427
QY 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
DB 1428 ACCGAAGCTTATCGCGGCTCTCTTATCTACTTGGCTGGAGCCTTCGACGAGCCCGGT 1487
QY 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
DB 1488 ATCAACGCTCCATTCATCGACAGTGATCGCAGCGCGGCTTCGTCAACTTCGACACGCT 1547
QY 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
DB 1548 CACTACTTCAGATTGATAGTGGCGCGAGCTGGCGCAGGCTGCTCGTTACTCAAAAAG 1607
QY 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
DB 1608 TGCACGCTGCATCGATTGTCACCGCGCGAGGCTCTCGGGGGTACCTCCCAACACCATC 1667
QY 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
DB 1668 CGCGGGGATCTAGATGCGACATTCGACATCTCCCTTCTTGAAATGATGAGCTCTGAAA 1727
QY 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
DB 1728 CGTGTGGCGGAGATAATGGCGCGAGAAATCCCAACAAAGAGGTCACTACTCTTCTTCCA 1787
QY 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
DB 1788 CAAGCTATCCAACTGGGATCGCAACGACCTTCTTACCCGTCGCGGACGCTTACCCTCA 1847
QY 541 AsnGlyValaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCys 560
DB 1848 AATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCCTAGATCGGGAGGTGAGTGC 1907
QY 561 PheProAsnProValPheProSerAspGlyLeuGluIleLeuAsnPheGluGlyAla 580
DB 1908 TTCCTTAAACCCGCTGTTCCTCAAGCATGACGGCTGGAACTAATCACTTTCGAGGGGCA 1967
QY 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1968 TGCCTTACATATGAGGAGAGATCAACAGCTCGGCTCAACGCTCGCATTTGGAGGCGAG 2027
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
DB 2028 ATGCTGGGCATCCATCTATCGTTTCGACGGTATCCAGGCCCTACTTCTCGAGAGACAATC 2087
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 2088 ACTGTACGAACACTTCAACGAGGCTGATCAGCTTCGCCCGGAGAAAGCACCTTTGAATTC 2147
QY 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
DB 2148 CGCTTATTCACCGGAGGTCATCAAACTTTTCAGGACGGGACATTTCTCATCGATGGA 2207
QY 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
```



```
Db      2208  GATATGTTCCGGTTGGTTTACACTGGCGTGGCAGCTCCAGGCT 2255
|||||
US-10-691-383-1
; Sequence 1, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR FILING DATE: 2003-10-21
; PRIOR FILING DATE: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-10-691-383-1
Alignment Scores:
Pred. No.: 0 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-691-383-2 (1-676) x US-10-691-383-1 (1-2931)
Qy      1  MetLeuCysHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20
Db      228  ATGCTTTGCCATGAGCGGACACACAGAGGCTCTCTATGCTGACACCGGAGTGCTT 287
Qy      21  ArgLeuLeuThrSerGluGlnArgAlaGlyGlyTrpArgArgGlnLeuGluGlyGly 40
Db      288  CGGTTGCTCATCAGACGACGCGCTAAAGTTGGAGACCCAGTTAGAGGGGGAGAAA 347
Qy      41  SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db      348  TCACTAGGTTTTCATCCAAGCGAGACGCTTATATCAAGTACTTGGAGGCTCTGAGACT 407
Qy      61  TrpLysLysValLysLeuProThrAspGlyLysSerAlaSerLysHisLeuGlyLys 80
Db      408  TGGAGAAGGTTAAGCTTCAACGAGCGGATATCGGCTTCCAAGATCCTGGGTAAAT 467
Qy      81  MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db      468  ATGGCCAGGGTCCGCACTGCTACCGCTTGCTGTGGTACTGGCGGACACCTGTTGGCA 527
Qy      101  PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGly 120
Db      528  TTCACGAGGTCACAGCCAGTGTTGTTTCCCTCAGGACACACAGCAGCGGGAGGGA 587
Qy      121  ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db      588  AGACACCTCCAGACTGTACAACTCCGACGATCGCTGGATCCGACGCGCCGCAATAGA 647
Qy      141  ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db      648  AGGGACACCGTAGCTTTTGGCTCGCGCGCGATGCCGCCAGCGGAGAACGTTGAGCGACA 707

161  GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
708  GGGACTGTCTGCCAAATCACTAACGGAGAACTGATTTGGCTACCATGTTCCACAAGTCT 767
Qy      181  LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
768  CTGCCACAGATGAACCTGGGACAGGTAAACCGACAGACTTCGCTATCTCGAGGACTGC 827
Qy      201  IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
828  ATCTTAAACGGAGATTTTCAGCATTTGCGAGACGCTGCTCGGGAGACCCGCGGCTCGC 887
Qy      221  LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
888  CTCGTCAATCTACCGCTGCGTTTGGCATATCCGCTCCGCGATTTCTCGGCTACG 947
Qy      241  ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
948  ACAATACCCCGGTACCTTACCTTCTCTGAGCTCGCGCTCAGTTGGCGAGGCTA 1007
Qy      261  TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
1008  TACTGGATGGCGCTGGCCAGGATGTACCTTTATGCAATATGGCAGCCGACGAAATTACC 1067
Qy      281  ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
1068  ACTACCGCGCGAGCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGACGCGCTGTCG 1127
Qy      301  IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
1128  ATAGGGTCCGATGGTACGGTGGACCGCTTCTCCAGCTTCTCCGAGCGACCTTCGTTGGT 1187
Qy      321  ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle 340
1188  GTTGAACAGCGGCGCTTTGTCTCTCAGCTGCTCGTGAACAGCTTCCACCATCGACGCTATT 1247
Qy      341  ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
1248  ACGGTCGAACCGAAGCAGGAGACATTCGCCGCCGACTTGAACATATATGGTCGATTTTGCAC 1307
Qy      361  GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
1308  GAATGGCTGAACATTCAGAATGGTGGACCCCGCGCGCGCTTCCGAGAGTTAGACGAAGAG 1367
Qy      381  LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
1368  CTGGGTTTATCCGTAAACCGCGCGGACCTGGCAGGAGGCTCTCTTCGTGGACAATATCAAC 1427
Qy      401  ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
1428  ACCGAAGCTTATCGCGGCTCTCTTATCTACTTGGCTGGAGGCTTTCAGCAGGCGCGCT 1487
Qy      421  IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
1488  ATCAACCGCTCATTCATCGACAGTATCGGCGCGCGGCTTCTCGTCAACTTCGGCACGCTCT 1547
Qy      441  HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
1548  CACTACTTTCAGATTGATAGGTGCGCGCGAGCTGGCGAGCGGTGCTCGTGTATCCAAAG 1607
Qy      461  TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
1608  TGGCAGGTGCTATGATTTGACGCGCGCGAGGCTCTCGGGGGTACCCCTCCACAACACCATC 1667
Qy      481  AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
1668  GCGGGGATCTAGATGAGACTTCCACATCTCCCTTCTTGAATAATGATGAGCTCTTGAAA 1727
Qy      501  ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnGlnValThrTyrLeuLeuPro 520
1728  CGTGTGGCGGAGATTAATGGCGCGAGATCCCAACACAGAGGTCACCTACTCTTCTTCCA 1787
Qy      521  GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
```

```
Db 1788 CAGCTATCCAGTGGGATGCCCAACGACCCCTCTACCGTCCGGCCAGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGlyCys 560
Db 1848 AATGAGCAATTGCGACAGTTCTGAGGCCCTCATTTGGCCTTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuLeuLeuLeuLeuLeuLeuLeu 580
Db 1908 TTCCCTTAACCCCGGTTCCTCAAGCGATGAGCGCTGGAATCAATCAACTTCGAAGGGCA 1967
Qy 581 CysLeuThrTyrgluGlyGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCTTACATATAGGAGAGATCAACAGCTCGCGTCAACGTTCGATTTGGAGGCAG 2027
Qy 601 MetLeuGlyLeuHisTyrgPheAspGlyLeuGlnGlyLeuLeuLeuGlyGluThrile 620
Db 2028 ATGCTGGGCATCCACTATCGTTTCAGCGTATCCAGGCCCTACTTCTCGGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
Db 2088 ACTGTAGAACACTTCACCGAGGAGCTGATGAGTTCCCGGAGGAAGCCACTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCAGGAGAGTCAATCAACTTTCCAGGACGGGACATTCCTCATGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGGTGGCGGACTGCCAGGCT 2255
```

RESULT 3

```
US-10-132-134-11
; Sequence 11, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-20US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 11
; LENGTH: 9579
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-11
```

```
Alignment Scores:
Pred. No.: 0.000459 Length: 9579
Score: 131.00 Matches: 152
Percent Similarity: 30.12% Conservative: 82
Best Local Similarity: 19.56% Mismatches: 254
Query Match: 3.71% Indels: 289
DB: 15 Gaps: 38
```

```
US-10-691-383-2 (1-676) x US-10-132-134-11 (1-9579)
```

```
Qy 5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
Db 6726 GCGCAGATGACGACGCCGCCCGCGACCTCGACCCGGGACGACTTCTACCGCGCTCGG 6785
Qy 23 uThrSerGluAlaArgAlaLysGlyTyArgArgGlnLeuGluGlyGlyLysSerLeuG1 43
Db 6786 CACGCGGGGCTCGACGCCCGCGGCTAGCGCGGACGCTCGAAGGGTTCACCGACTCGA 6845
Qy 43 yPheHis-----ProSer-----GluThrProTyrlleLy 53
Db 6846 CGCCACCGGCTGCTCGTACGGGTGCGCGAACCAGGCGCATGTGCCAGGACCCGACAGCA 6905
```

```
Qy 53 sTyrLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72
Db 6906 GCACGTGCAT---CTCCCGCCCTGGCGCTCGTGGGCTG---ACCAGGAGTGTTCAGCA 6959
Qy 73 -----AlaSerLysIleLeuGlyLysIleMe 81
Db 6960 CGCGTGGGGCGGGCCGACGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7019
Qy 81 tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101
Db 7020 GGAGCGCACCCCG---GGCATCTGCTGCGCGG----- 7050
Qy 101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
Db 7051 -----ACGTCCGACGCCCGTCTTC-----CATCGGCTTTCCTCGA 7085
Qy 119 ----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
Db 7086 CGAGGACGGCGCGTCTGGCGCGGTCGAGGACGCCGAGTTCACCGCGGCGGACCTTGA 7145
Qy 134 pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147
Db 7146 GCGGCACTCCCGGTGAGCGCGACGCACTCTGTCGCACTGCCGAGCGTGCCTGCC 7205
Qy 148 -----SerArgArgAspAlaAlaArgArgG1 156
Db 7206 GGTGCTGGAGACCGCGTGTGTACGGGGGAGTGGCAGCAGTCGAGGAGCCGTGCGCGGA 7265
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169
Db 7266 GCGGAGCGCTCGTGACCGTTCGGCGGTCGCGGACGGCGCGCGCTCGTGGCGTC 7325
Qy 170 -----GluThrAspLeuAlaTh 175
Db 7326 GCTGCGGACGCGTCCCGACCTGCTCAAGTTTCGACCTCGCGGACATCGACCTCGAC 7385
Qy 175 rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194
Db 7386 GCATCTCCACGCGTACGGCTTCGAGTCCATCGCGCTGGCCAACTGGCCCTCGGAATCAA 7445
Qy 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214
Db 7446 CGCGTCTCGGACGACCTCACCCCGCGCTCTTCGAGTGTCTCGAC----- 7497
Qy 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerG1 234
Db 7498 -----ATCCGACGCTCGCGAGTACTGCTCGACCGCTACGG 7535
Qy 234 yProAlaPheSerAlaThrThr-----IleProProValProThrLeuSe 249
Db 7536 CCCGAGCTGAGCTCCCGACGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 7595
Qy 249 rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspVa 269
Db 7596 GTCCCA-----GT 7604
Qy 269 lProPheMetGlnTyrglyThrAspGluIleThrThrAlaAlaAlaAlaAlaG1 289
Db 7605 GCGATGCGCGCACCCCGCGCGGACGAC-----GACGCGGTGGCCATCGTCCGCGC 7655
Qy 289 yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309
Db 7656 TGCGGACGCTTCCCGCGCGGAC-----GACCTGGACAC 7691
Qy 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321
Db 7692 CTTCTGCGACAGCTGCGCGCGGCGGAGGACCTGATCGCGCACTACCCCGCGGACCGCT 7751
Qy 321 lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleTh 341
Db 7752 CGACGGGGGCCCCCTACGCGGAGGTGCTGCGCGCGGCGGAGTTCCTCCGAAGTTTCCGCGCG 7811
```

```
Qy 341 rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspG1 361
Db 7812 GATCGAG---GGCGTGGACCGCTTCGACCGCGACTTCTTCACCTGTCGGCGTGGAGC 7868
Qy 361 u-----TrpLeuAsnIleGlnAs 367
Db 7869 GGAGCTGATGACCGCGAGACCGGCTGGCCCTGGAGACCGTGTGGCGCGCTGGAGAA 7928
Qy 367 nGlyGly-----ProPheAlaGlyProGluGluLeu-----As 378
Db 7929 CGGCGGCTACGCCCGCGCGCCCTCCCGAGAACACCGGGTCTACTTCGGCGCTCCGG 7988
Qy 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391
Db 7989 CAGCGACTACCAACCTGCTCAACGCGCGTGGCGTGGCACCAGCGCTTTCACCGGCAC 8048
Qy 392 -----ArgValSerPheValAspAsnIleAsn----- 400
Db 8049 CGGCAACGCCCACTCGATGTGGCGAACCGGATCTCTACGTCTGTGACGTGACGGGCC 8108
Qy 401 -----ThrG1 402
Db 8109 GAGCGAACCCGTCGACACGGCTGCTCCAGCTCGCTCGCGCTGACCGCGCGTGA 8168
Qy 402 uAlaTyrArg-----GlySerLeuIleLeuLeuG1 412
Db 8169 GCACATCGGTCGGCGCGGATCGAGATGCGCATCGCGCGGCTGTCAACCTGCTGCTGAG 8228
Qy 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerArgGlnAl 432
Db 8229 CGTGAGACACCTTCGCCCGGACGCATGCGCGGC---ATGCTCAGCCCGACGCGCGT 8285
Qy 432 aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451
Db 8286 CAAGACCTTCCTCCCGCGCGGCGGCGGCTAGCTCGCTCCGAGGCGCTCGCGCGTGT 8345
Qy 451 uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPh 466
Db 8346 GCTCAAGCGGCTCGCCCGGCGGACGCGGACGCGCATCTCGGCGGCTGCTCGCGGG 8405
Qy 466 eAlaArg----- 468
Db 8406 CAGCGCGGAGAACACCGCGCGCGCGCGGTTGCTGACCGCCCGAACGGAAGGCGCA 8465
Qy 469 -----ProGluAlaLeuGlyGlyThr-- 475
Db 8466 GGCGCGCTGATCCAGACGCCATCGCGCGGCTGACCGCGACGATCGGCTACGTCGA 8525
Qy 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuGluAs 495
Db 8526 GGCGCAGCGCACCGGCGCACCGGC---CTGGCGGACCGCGTCAAGGTCAACGCGCTCGACAG 8582
Qy 495 nAspGluLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515
Db 8583 CGCTACCGCGCGCTGCGCACCGCGAG----- 8610
Qy 515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535
Db 8611 -----GGCGGCGCGCGCGCACCGCGCGCGCGCGCGCGT 8639
Qy 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaIath 546
Db 8640 CGCGCTCGGCTCGGTGAAGACCAACATCGGCGACGCGGAGTCGCGCGCGCGCTGGCGG 8699
Qy 546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlyCysPhePro----- 562
Db 8700 AGTCTGAAGGTGTGCTCGCCATG---CGTCACCGGAGGTGCGCGCGCGCTTGCATG 8756
Qy 563 -----AsnProValPheProSerAspGlyLeuLeuIle 575
Db 8757 CGACCGGCTCAACCGGACCTGCGCTCGACCGCGGATTCGAGGTGCTGA 8805
```

RESULT 4

```
US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staiffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-20S
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 52101
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

Alignment Scores:
Pred. No.: 0.00725 Length: 52101
Score: 131.00 Matches: 152
Percent Similarity: 30.12% Conservatives: 82
Best Local Similarity: 19.56% Mismatches: 254
Query Match: 3.71% Indels: 289
DB: 15 Gaps: 38
```

```
US-10-691-383-2 (1-676) x US-10-132-134-1 (1-52101)
Qy 5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
Db 14734 GCAGCATGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14793
Qy 23 uThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyGluLysSerLeuG1 43
Db 14794 CACCGCGGGCTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14853
Qy 43 yPheHis-----ProSer-----GluThrProTyrIleLe 53
Db 14854 CGCCACCGGCTGCTCGTACGGTTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCA 14913
Qy 53 sTyrLeuGluGlySerGluThrTyrLysLysValLeuLeuProThrAspGlyIleSer-- 72
Db 14914 GCAGTGCAT---CTCCCGCGCTGGCGCGCTCGTGGGCTG---ACCCAGGCTGTTCA 14967
Qy 73 -----AlaSerLysIleLeuGlyLysIleLe 81
Db 14968 CGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 15027
Qy 81 tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101
Db 15028 GGAGCGCACCGG-----GCATCGTCTGGCGG----- 15058
Qy 101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
Db 15059 -----ACGTCCGACCGCGCTCTTC-----CATCGGCTTCTCTCGA 15093
Qy 119 ----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
Db 15094 CGAGGACGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15153
Qy 134 pProThrAlaProAsnArgAspAsnValAlaPheAla----- 147
Db 15154 GCCGCACTCCCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15213
Qy 148 -----SerArgArgAspAlaAlaArgArgG1 156
Db 15214 GGTGCTGGAGACGCGCGTGTGTACGGGGAGGTGGCAGCAGTCCGAGCGCGCGCGGA 15273
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169
Db 15274 GGCGGACCGCTCGGTGACCGTTCGCGCGGTTCGCGGACGGGCGCGCGCGCTGCTGCGG 15333
```

```
QY 170 -----GluThrAspLeuAlaTh 175
Db 15334 GCTCGGGAGACCGTCCGCGACCTGCTCAAGTTCCGAGCGGACATCGACTCGACAC 15393
QY 175 rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194
Db 15394 GCATCTCCACGCGTACGGCTTCGAGTCCATCGCGCTGGCCAACTGGCCTCGGAACCTCAA 15453
QY 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214
Db 15454 CGCGCTCTCGGACGCGACCTCACCCCGCGCGTCTTCTCGAGTGTCCGAC----- 15505
QY 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerG1 234
Db 15506 -----ATCCGAGCTCGCGGAGTACTGTCTCGACCGCTACGG 15543
QY 234 yProAlaPheSerAlaThrThr-----IleProProValProThrLeuSe 249
Db 15544 CCCGAGCTGAGCTCCCGAGGCGCGAGCGCGCGCGCGTCCGCCGCCCGCGCC 15603
QY 249 rSerProGluLeuAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspVa 269
Db 15604 GTCCCCA-----GT 15612
QY 269 lProPheMetGlnTyTrpGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaG1 289
Db 15613 GCCGATCGCGCACCGCGCGCGGAGCAGC-----GACCGGTGGCCATGTCTCGCGCGC 15663
QY 289 yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309
Db 15664 TGCCGGACGGTTCGCCGCGCGGAC-----GACCTGGACAC 15699
QY 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321
Db 15700 CTTCTGCGCAGCAGTGGCGCGCGGCGAGGACCTGATCGCGGACTACCCCGCGCAGCGCTT 15759
QY 321 lGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleTh 341
Db 15760 CGACGGGGCCCCCTACCGGAGGTGTCTCGCGCGCGGACTTCCGAGGTTTCGCGCGCG 15819
QY 341 rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyTrpMetValAspPheAspG1 361
Db 15820 GATCGAG---GGCGTGGACCGCTTCGACGCGGACTTCTTCCACCTGTGCGGTGGAGGC 15876
QY 361 u-----TrpLeuAsnIleGlnAs 367
Db 15877 GGAGCTGATGACCCCGCAGCACCGGCTGGCCCTGGAGACCGTGTGGCGCGCTGGAGAA 15936
QY 367 nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378
Db 15937 CGCGCGGTACGCCCGCGCGCGCTCCCGGAACACCGCGGTCTACTTCGCGGTCTCCGG 15996
QY 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391
Db 15997 CAGCGACTACCACTGCTCAACGCGAGTGGCGTGGCCACCGCGCTTCCACCGCCAC 16056
QY 392 -----ArgValSerPheValAspAsnIleAsn----- 400
Db 16057 CGGCAACGCCCATCTGATGTGTGGCAACCGGATCTCTCTGCTGACGTGACGCGGCC 16116
QY 401 -----ThrG1 402
Db 16117 GAGGAAACCGGTGCACAGCGGCTGCTCGCTGCTGCTGCGGTGACCGCGCGCTCGA 16176
QY 402 uAlaTyArg-----GlySerLeuIleLeuLeuG1 412
Db 16177 GCACATCGGTTCGGCGCGATCGGAGATGGCCATCGCGCGGTGTCTCAACCTGTGCTGAG 16236
QY 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432
Db 16237 CGTGGACACCTTTCGCCGCGACGCATGTGGCGGC---ATGCTCAGCCCGCAGCGCGCTG 16293
```

```
QY 432 aGlyPheValAsnPheGlyThrSerHisTyTrpPheArgLeuIleGly---AlaAlaGluLe 451
Db 16294 CAAGACCTTCTCCCGCGCGCGGACGCTACGTCGCTCGAGGGCGTCCGCGGTGCT 16353
QY 451 uAlaGlnArgAlaSerCysTyTrpGlnLys-----TrpGlnValHisArgPh 466
Db 16354 GCTCAAGCCGCTCGCCCGCGCGAGCGGACCGCGCATCTGCGGGCGTCTGTCGCGGG 16413
QY 466 eAlaArg----- 468
Db 16414 CAGCGCGAGAACACACGCGCGCGCGCGCTTCTGCTGACCGCCCCAACGCGAGGCGCA 16473
QY 469 -----ProGluAlaLeuGlyGlyThr-- 475
Db 16474 GCGCGCCCTCATCAGGACGCCATCGCGGCATCGACCGGACGATCGCTACGTCGA 16533
QY 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuAs 495
Db 16534 GCGCACGCGCACGCGGACCGCGC---CTGGCGGACCGCGTCTGAGGTCAACGCCCTCGACAG 16590
QY 495 nAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515
Db 16591 CGCTACCGCGCGCTCGCACCGCGCGG----- 16618
QY 515 lThrTyTrpLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyProSe 535
Db 16619 -----GGCGGCGCGCGCACGCGCGCGCGCTG 16647
QY 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
Db 16648 CGCGCTCGGTGCGTGGAAGACCAACATCGGCCACGCGGAGTCCGCGCGCGCTGGCGCG 16707
QY 546 rValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562
Db 16708 AGTGTGAAGTGTCTGCTCGCCATG---CGTACCGGAGCTCGCGCGCGCTTGCACGTG 16764
QY 563 -----AsnProValPheProSerAspAspGlyLeuLeuLeuIle 575
Db 16765 CGACCGGCTCAACCCGACCTCGCTCGACGCGGATTCGAGGTCTGTA 16813
RESULT 5
US-10-437-963-29553
; Sequence 29553, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 29553
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34043C.1
US-10-437-963-29553
Alignment Scores:
Pred. No.: 2,91e-05 Length: 1236
Score: 129.00 Matches: 110
Percent Similarity: 33.83% Conservative: 50
Best Local Similarity: 23.26% Mismatches: 142
Query Match: 3.66% Indels: 172
```



```
Qy 63 Lys-----ValLysLeuProThrAspGlyLeuSerAlaSerLysLeuGlyLys 79
Db 877 GAGATTTCGGCTCTTCGAACCCACGAGCCACGAGCTCCCGCGTCTACGAGAAG 936
Qy 80 IleMetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeu 99
Db 937 CTCTACGACACCATCCGCGAGCGCGCC----- 966
Qy 100 AlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGlu 119
Db 967 -----GACTCCCGCGTCTCCGAGCGGTCTTC---GAGTGGCGACCCGCGTGGCGCGC 1017
Qy 120 GlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsn 139
Db 1018 GCCGCCAC-----ACACCGACACCCCGGATCGGCTGGCGCGCACAC 1065
Qy 140 Arg----ArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgGluArgAsp 158
Db 1066 GCCCTCGCGACACGCTCGTGTCTCGACATCCGAGCGCTTGGGGGAGATCGAG 1125
Qy 159 -----GlyThrGlyValCysGlnIleThrAsnGlyGluThrAspLeuAla 174
Db 1126 TTTTTCATCTCCGCGCGCTCGCTG-----TCCCGGAGCTGTGT 1167
Qy 175 ThrMetPheHis---LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp 193
Db 1168 CGCTGTTCACGCATGGGGTCCCC----- 1194
Qy 194 PheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValPro 213
Db 1195 -----ATCCTCGAAGGC---TACGGCTTCACGAGACGAGCCGCG 1230
Qy 214 AlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSer 233
Db 1231 GTG-----CTCGGTCAACCCCTACGAGGACCCCATCTCGCGCACCATC 1275
Qy 234 GlyProAlaPheSerAlaThrIleProValProThrLeuSerSerProGlu--- 252
Db 1276 GGCCCGCGGTACGGACACCGAGCTACCGTCGAGACCATCGCAGCCCGCGAGCAA 1335
Qy 253 -----LeuAlaAlaGlnLeuAlaGlu 259
Db 1336 CGCCAGCGCTGTGACGGCGCGCGGAGCTGCTCGCGCGCGCCACAAAGTGTTCGAC 1395
Qy 260 LeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIle 279
Db 1396 GGCTACTGG-----GGGCTGCCCGATGCC 1419
Qy 280 ThrThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaVal 299
Db 1420 ACCGACCGCGGTCTCATCCCGGAGGGCAGAGAGTGTTCGACACCGCGGAGCTGTC 1479
Qy 300 SerIleGlySerAspGlyThrVal-----AspProPheSerGlnLeuPheArg--- 315
Db 1480 GAATCTCGCCCGACGGCTAGCTCGGTCTCTCGAAGCGCGCAAGCAACTGCTGACCTC 1539
Qy 316 AlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPhe 335
Db 1540 TCGACGGGCAAGACGTGCGCCCGCCCATCGAGGACGCGTTTCGCGCGCGCGCGCTG 1599
Qy 336 ThrIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyr 355
Db 1600 GTGACAGCGCGGTCTGTCGGGAGCGGCAGAGTGTCTCTCCGCGATCCTG----- 1653
Qy 356 MetValAspPheAsp-----GluTrpLeuAsnIleGlnAsnGlyGlyProProAla 372
Db 1654 GTTCCGAACCTTCGACGCGGTCTCGAGTGGGCGCCCGACGAGAGATCGGCTCCCC--- 1710
Qy 373 GlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArg 392
Db 1711 -----GACGACCGCGAGCGCATCTGCGCGAGCGAGCGCTCGCGCGCGC 1755
Qy 393 Val---SerPheValAspAsnIleAsnThr-----GluAlaTyrArg-GlySerLeuI 409
```

```
Db 1756 ATCAGTTCGGCGTTCGACGAGCTGCACACCGGCTTCGAGCCTACGAGCGGATCA----- 1810
Qy 409 eLeuLeuGluLeuGlyAlaPheSerArgPro 419
Db 1811 -----AGCAGTTCGGCGCTCTGTCGCGACCG 1834

RESULT 7
US-10-329-079-48
; Sequence 48, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFPA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 7185
; TYPE: DNA
; ORGANISM: Streptomyces refuineus
US-10-329-079-48

Alignment Scores:
Pred. No.: 0.00186 Length: 7185
Score: 124.50 Matches: 170
Percent Similarity: 29.79% Conservative: 60
Best Local Similarity: 22.02% Mismatches: 240
Query Match: 3.53% Indels: 302
DB: 15 Gaps: 38

US-10-691-383-2 (1-676) x US-10-329-079-48 (1-7185)
Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMet----- 14
Db 3493 CTGCGGCGCCACCGCTTCGACCTGGAGCGCCACCCCATCGGGCTGGCTGCTGGAG 3552
Qy 15 -----ProAspThrGlyValLeuArgLeuLeuThrSerGluGlnArgAlaLysGlyTrp 32
Db 3553 ACCGGCGCGCACGCGGGTCTCGTGTGCTGTGTCGACCATCGCAGCGAGCGGTGG 3612
Qy 33 -----ArgArgGlnLeu-----GluGlyGluLys 40
Db 3613 TCGGCGAGGAGGCTCTCTCGCGACCTGTTACCGCCTTACCGCACCGCGCGGCGCGG 3672
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 3673 GCCCGNACTGGCGGCGCTGCGCGGTGTCAGTAGTGACTAC-----GCCCTG 3720
Qy 61 TrpLysLysValLysLeu-----ProThrAspGlyLysSerAlaSerLysIleLeuGly 78
Db 3721 TGGCAGCGGCGGTCTCTCGCGCATCCCGCGACCCCGCGCAGCAC----- 3765
Qy 79 LysIleMetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCys 98
Db 3766 -----GCAGCGCGCCAG 3777
Qy 99 LeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly 118
Db 3778 CTGAGTACTGGGAGCGGAGCTGCGCGGTCTG---CCGAGGAGCTGAGGCTCCCGGCC 3834
Qy 119 Glu-GlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-Alap 138
Db 3835 GACCGGCGCG-----GTCCGCGCGCTCCCGTCC 3861
Qy 138 roAsnArgArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArg 158
Db 3862 CGCACCGCGCGCGAGGTCTGCTGCTGACGCTCCCGCATCTCCGTCCACACCGCGGTGTCGAC 3921
```


Qy 158 spGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheH 178
Db 3922 CTGGCCCGGAGCGTGGCGGAGCGTGTTCATGCTCGTCCAGGCGCG-----C 3969
Qy 178 isLysSerLeuProHis---Asp-GluLeuGlyGlnValThrAlaAspPheAlaIle 196
Db 3970 GTCGACGCTTCTTCACCCCGATGGCGCGCG----- 4002
Qy 197 LeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAsp 216
Db 4003 -----GAGGACATCCCGTCGGCACC 4023
Qy 217 Pro---AlaGlyArg-----LeuValAsnPro 224
Db 4024 CCGGTGGCGCGCGCACCGACAGCGCGGTGGAGACCTGTGCGATTCTTCGTCAAC--- 4080
Qy 225 ThrAlaAlaPheAlaIleAspIleSerGly---ProAlaPheSerAlaThrThrIlePro 243
Db 4081 ACCCTGTCTCTGGCGGACGACCTCCGCGCACCCCGGTTCGCC----- 4125
Qy 244 ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMet 263
Db 4126 -----GAGCTGTGGACGGGTCCGCGAGACCGCGCTGGCC 4161
Qy 264 AlaLeuAla---ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThr 282
Db 4162 GCCTACGCGCACGACGACCTGCGCTCCGCTCCGAGCAG-----CTGGTGGAG 4203
Qy 283 AlaAlaAlaLeuAlaGlyMetGlyGlyPhePro----- 294
Db 4204 CGCTGAGCCCGCGCGCTGCTCGCGCGCACCGCTCTTCAGTGGCGCTCTCTCTGC 4263
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
Db 4264 AACAAACCGGAGGAGCAGCTGGCGCGCAGCGCTCCCGCGCCCC----- 4308
Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSer 334
Db 4308 ----- 4308
Qy 335 PheThrIleAspAlaIleThrValGluProLysGln---GluThrPheAlaProAspLeu 353
Db 4309 -----GGACTGCGCGTCAAGCCCCACACGAGTGGAGACCGCGGTGCAAGTTC 4356
Qy 354 AsnTyrMetValAspPheAspGlyTrpLeuAsnIleGlnAsnGlyGlyProProAlaGly 373
Db 4357 GACCTGATGTTACCTTCTCTGGAG-----GGCCACCGGAGGAGCGCGCGCGCGG 4410
Qy 374 ProGlu-----GluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeu 390
Db 4411 ATCGAGACCGCGCTGGAGTACAGCGCGCACCTCTTCGACAGGAGACCGCGGACCTG 4470
Qy 391 AlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeu 410
Db 4471 -----CTGAGGCGGTTCGCGCGGATGCTCGCGCTC 4500
Qy 411 LeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArg 430
Db 4501 TGGGCG-----GCGACCGGCG-----GGCCCC----- 4524
Qy 431 GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu 450
Db 4525 -----ATCGGAGCGCGGAG 4539
Qy 451 LeuAla-----GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 467
Db 4540 CTGCTCGCGCGCGACGAGCGGCACACGCTGTGCGCGAGTGGAAACGCGCACCGCGCGCG 4599
Qy 468 ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAlaAsp 487
Db 4600 -----GGCTGTGTGGCGACGCTCCGCGAGATGTTCCAGGAGCAGGTGCGCGCGACT 4650
Qy 488 PheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAla 507

Db 4651 CCCGACCCCGCGCGTGGAGCAGCGCGCGCGCGGCTGACGTACGCGCACTCAACGCC 4710
Qy 508 AlaGlnAsn-----ProAsnAsnGluVal 515
Db 4711 CGGCGCAACCGGTCGCCAGGTGCTGTCCGACCGGCTCGGCGCGCGCGGTG 4770
Qy 516 ThrTyrLeuLeuProGlnAlaIle---GlnValGlySer----- 527
Db 4771 GCCGTGTGATGCGCGCTCCCTCGGCGAGTCAACCGCTGTGCGGTGCTCAAGGCC 4830
Qy 528 -----ProThrHisProSerTyrProSerGlyHisAlaThr 539
Db 4831 GCGCGCGCTACGTGCGGTGGACCCCGCGCACCGGAGGACGATCGCTTCATGCTG 4890
Qy 540 GlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu 559
Db 4891 CGGACAGCGCGCGCGCTGTCTGCGCGCGAGTCTGCGCGCGGAGCGCGGAG 4950
Qy 560 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly 579
Db 4951 ATCGCGCGGTTCGCGTCTGTGTCGCGACGACGCGCGCGCGCGCGGACGCGCGG 5010
Qy 580 -----AlaCysLeuThrTyrGluGlyGluIleAsnLys-----LeuAlaValAsn 594
Db 5011 CCCTGCGCGCGCGACCTCACC---GACGCGGACCGGAAACGCCCTTGACCGCGCGCAAC 5067
Qy 595 ValAlaPhe-----PheAspGly----- 597
Db 5068 GCGCGTACGTGCTCTACACTCGGCTCCACGGCGCGCGCGCGGAGGCGGTGACCGAG 5127
Qy 598 -----GlyArgGlnMetLeuGlyIleHisTyrArg----- 607
Db 5128 CACCGCGTCTGTGCTGCGTGGCGTGGCGCAGCGGAGCGGTACCGCGTGGCGCGCGC 5187
Qy 608 -----PheAspGly----- 610
Db 5188 AGCGGTGTGTCAGTCTGCTGCGCTGCTTCGACGCGCGCGGTCTGAGTCTGATG 5247
Qy 611 -----IleGlnGlyLeuLeu 615
Db 5248 GCGTTCGCCACCGGAGGACGCTGCTGCTGCGCGCGCGCGCGGAGGAGTGTG 5307
Qy 616 LeuGlyGluThrIleThrValArgThrLeu 625
Db 5308 CTCGGCGAGACCATCGCGCGCGCGGATC 5337

RESULT 8

US-10-329-079-34
; Sequence 34, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 61944
; TYPE: DNA
; ORGANISM: Streptomyces refuineus
US-10-329-079-34
Alignment Scores:
Pred. No.: 0.0623 Length: 61944
Score: 124.50 Matches: 170
Percent Similarity: 29.79% Conservative: 60
Best Local Similarity: 22.02% Mismatches: 240
Query Match: 3.53% Indels:

Db 51201 CACCGCGTCTGCTGCTGCGCGCGAGCGGAGCGGTACCCGGTGGCGGCGGC 51260
Qy 608 -----PheAspGly----- 610
Db 51261 AGCGGTGCTGCGAGCTGCGCTGCGCGCTTCGACGCGCGGCTGGAGCTGCTGATG 51320
Qy 611 -----IleGlnGlyLeuLeu 615
Db 51321 GCGTTCGCCACCGAGGAGCGCTGCTGCTGCCGACCGCGCGCTCTTGGCGGGGAGCTG 51380
Qy 616 LeuGlnGlyThrIleThrValatgThrLeu 625
Db 51381 CTCGGCGAGACCATCGCGCGCGCGGATC 51410

RESULT 9

US-10-282-122A-31609
; Sequence 31609, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31609
; LENGTH: 18930
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-31609

Alignment Scores:

Pred. No.:	0.0104	Length:	18930
Score:	124.00	Matches:	147
Percent Similarity:	30.99%	Conservative:	56
Best Local Similarity:	22.44%	Mismatches:	226
Query Match:	3.51%	Indels:	229
DB:	13	Gaps:	31

US-10-691-383-2 (1-676) x US-10-282-122A-31609 (1-18930)

Qy 111 ProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCysThrAsn----- 128
Db 573 CTTGATCGGAGCGGCACAGTGGTGGCGGCGCACTTCAACGTGAGCTGTAATGGCGC 632
Qy 129 -----SerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146
Db 633 GCAGATCAACAGCGGAAACCTTGTATGTCACCTGACTGACGCGCGCGGTATGTCG-- 690
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
Db 691 -----GCGCTGGCGCTGCTCACTGTCACCGGACGCACTGCACCACTG----- 732
Qy 166 eThrAsnGlyGluThrAspLeuAla-----ThrMetPheHisLysSe 180
Db 733 -----GCGCGCACCGACCTGGCTATCAATGAGAGGGTAACACCTTACCGTCTGTCG 785
Qy 180 rLeuProHisAspGlu-----LeuGlyGlnValTh 190
Db 786 CGAGCCCGGCTCGACGGTTTCCTGGAGCGGAGTGGCGTATTGCTGGGCATCGCGGT 845
Qy 190 rAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG 210
Db 846 GGCCTG-----GCGAGCGGCAGTTTCAGCATACCT 878
Qy 210 uAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
Db 879 GCAACCGCCACAAAGCGAC-----GGCCAGGCGCTGGAGATCAGCGCGCAGCGCGC 932
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProProVal-----ProThrLe 248
Db 933 TGGCAATATCTCGCCAGTTGCCAGCATCACGGCG-----CCGAGCGCTGACCAACCCAGATAC 989
Qy 248 uSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTy-TrpMetAlaLeuAlaArgAs 268
Db 990 CACAGCCCTGAACACCGGACCGACCTG-----GCCCTGGCGCGATGG 1031
Qy 268 pValProPheMetGlnTyGlyThrAspGluIleThrThr-----ThrAlaAlaA 285
Db 1032 CGTCACCTTTCACCGGCGCGGTGGTGTCTACCGTGCAGGTGGCGGATGCTGCGG 1091
Qy 285 aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspG 305
Db 1092 TAACTCTGATTGGCACCCGT-----GTCTGCGTGGCGGATGG 1127
Qy 305 yThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr 325
Db 1128 C-----CTGTTACGCTCACCTCTGCGCTGCGCTGCCAGAC----- 1161
Qy 325 oPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy 345
Db 1162 -----AATGGTGAAGCACCTCGACGTGCGTCTGTTGGATGCGCG 1199
Qy 345 sGlnGluThrPheAlaPro-----AspLeuAsnTyMetValAspPheAspGluTr 362
Db 1200 CGGCAACAGCTCTGCCCATTTGCAATTCGATGCGCGGACATCACACCGCTGACGCGAGT 1259
Qy 362 pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr 382
Db 1260 AACCAATATTACAGTTGGTCTGCTGACGG-CCTGGCCCTGAGCGGTGGCGGCGGAGCCAGTG 1318
Qy 382 gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs 397
Db 1319 CCACTGTGGAAGTGGCGGATGCCAATGGCATGTGATCG-GCACGGGTGCTGCTGGGTGCC 1377
Qy 397 pAsnIleAsnThrGluAlaTyArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSe 417
Db 1378 -AAC-----GGCACCTTCTCTGATGACCTGCCCGCCCGCGCGC 1412
Qy 417 rArgProGlyIleAsnGlyProPheIleAspSerArgGlnAlaGlyPheValAsnPh 437
Db 1413 ACAGCCGGCGAGCAGTTGAGCTGTGGTGGACAGCGAC---CCAGCGGTAAATGCTTCACT 1469

```
Qy 437 eGlyThrSerHisTyrPheArgLeuIleGlyAlaAaglu----- 450
Db 1470 GGCTACCGAGTACGATGATACACTGACACCGGCCCGACAGCCCGAGCAATCTGGCCAT 1529
Qy 451 -----LeuAlaGlnArgAlaSerCysTyrGlnIlyStrpGlnValHi 464
Db 1530 CGATGCTGATGGCACCACCCTCACCGGTACCGCGCGGCTGGCAGCGGTGGAGGTGCA 1589
Qy 464 sArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLe 484
Db 1590 T-----GACGCCAACCGGCACGCTGATTGGCAGTCCCATTCGCCACGC 1631
Qy 484 uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuIlyAr 501
Db 1632 CGATGGCAGCTTACGATCGAAGTCCGACCAAGCCCAAGCCGCACTGCTGGACGT 1691
Qy 501 gValAla----- 503
Db 1692 GGTGGCCATCGATGACGGCGCGGTTCTCGTCCGCGCACAGATCACCGCGCCGACAT 1751
Qy 504 -----GluIleAsnAl 507
Db 1752 CACGCGCCCTCGCCGACCGCAGCACTGGCGGTGAGCGCGGACGCGTATACCGG 1811
Qy 507 aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlyse 527
Db 1812 CCGTGCCGAAACCGGGACGACCGTGGCGCATCGTGGTCCGACGGAACCGAATCGGCGAC 1871
Qy 527 r-----ProThr-----HisProSerTyrProSerGI 536
Db 1872 TGGCGTGTGGCGCGCGCGGTTCAGCCTCAACCTCAACCGCCACCAAGTTGACGG 1931
Qy 536 yHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIlyAlaLeuIleGlyLeuAspAr 556
Db 1932 CGAAGTCTCGAA-----GCCACGGCGACC-----GATGC 1961
Qy 556 gGlyGlyGluCysPheProAsnProValPhe-----ProSerAspAspGlyLeuGlu-- 573
Db 1962 CGCTGGCAACATCGCCAGCAGCGCGGATACCGCACCGGATACGACGGGTGCGATAC 2021
Qy 574 -----LeuIleAsnPheGluGlyAlaCysLeuThrTy 584
Db 2022 CACGCCACCGCAGCCCGACTGACTTGGTGTGATCGGCTCGCCGCGACCACTCAGTGG 2081
Qy 584 rGluGlyGluIleAsnIlyLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyI 604
Db 2082 TCGAGGCGCAA----- 2091
Qy 604 eHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgTh 624
Db 2092 -----GCCGGCTCTACCGTGCAGGTACGTGA 2117
Qy 624 rLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPheTh 644
Db 2118 CGCGGCGGCAACATC-----CTCGCGAC 2141
Qy 644 rGlyGluValIleIlyLeuPheGlnAspGlyThrPheSerIle 658
Db 2142 CGGACCGCTCGCC-----GCCGATGGCACCTTCACGTGC 2175
```

RESULT 10

```
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
```

```
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1
Alignment Scores:
Pred. No.: 0.0985 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 14 Gaps: 26
```

US-10-691-383-2 (1-676) x US-10-014-717-1 (1-68750)

```
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGCTGTGGGACGCGCGCTGGACGCGTGGACGCGCCAGCGCGAGCGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGCTCGAC-----CTGGTGCTCGGACGCTTCCATGCGAAGTGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGGAGCGCTCGCGGAGCGGTACATCATCGGCACCTCCGCATATGGAACGCTTCTGCG 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCy 126
Db 51681 GCT-----GCTGGAGAGCGTCCACAGATAGACGAGTTG 51713
Qy 126 sThrAsnSerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146
Db 51714 CTGCTCAGGCTTCAATCTCTGCTGTACAGGAGGTCAAGCGCATGGATGGAAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
Db 51774 CTGTGCGGA-----TCGGCATCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCGCTGCGCGAGCGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGAGGGGTGTTCCCGCACCTCCAGTCCTATTGAGTGTGTCAAGTTTCCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGCGGAGTATTGACCGGTAAAGCGCTCGCGCTCGAGATCTCTCTCCCTGGTGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTCGATATGGCGAGCGCAATCTATCGAGATTCGCCCATCGCCCTTACTCGAACGCGAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
Db 52049 CGTGGCGGCTGCTCGAGTCCGCGCGCGGGTGGTAGCACCGCTCGGGAATGTTCCAGAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGGAGATCGGACGAGCGGCGGACCGCGCGCGTCTCCCGGTGTGTCTGCC 52168
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGACGAGTACCATTTCCAGATGTTTCTCGCGCTCTCTCTCTCGCGGAGCA 52228
```

```
QY 267 -----ArgaspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284
Db 52229 AAGATTTTCGAGATTTCATCTCTGAAGTAGTGGCATTTCGATTCGACAGGAGCCAGC 52288
QY 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGGATACACATACAGAGTTTGACGTCTGTCGCGCCCAATGTC---ATCCA 52345
QY 304 pGlyThrValAspProPheSerGlnLeuPheAlaThrPheValGlyValGluThrG1 324
Db 52346 TGCACCCCGCAT-----ATAAGAGCCAGC----- 52370
QY 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGAAGCGTCTCTCTG----- 52385
QY 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTTGCTCGCGCCGAGGCCTTCTGCTGCTGTCGAGGGCACAGGGCATCC 52438
QY 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluG1 380
Db 52439 GATCTGTTTGATATACACAG---GGATTGATTGACGGGTGGCAGAGTACGAGATGA 52495
QY 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATACCATCCGCTCCTGCTCTGCTGCGACCTGTGTGACGTCTCTGCGCGGGT 52555
QY 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGGCTTTTCGCGAGCCGCTGATCTGCCAGGCGAGCGATCTCCGCGGGGATCTCTCGGACA 52615
QY 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427
Db 52616 GCAGTGATCTCTCG-----CGCGCCCGGGCATACGAGAGCGCGTGTGTA 52663
QY 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleG1 447
Db 52664 CAGCTCCGGTGAGTCG----- 52679
QY 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462
Db 52680 -GCGACCGAATTCGCGCGCCGCGTGCAGTACGCGCAGGAATGGCCGATGCTCCGCTGA 52738
QY 463 ----ValHisArgPheAla----- 467
Db 52739 CGTCTCATCGATGCGTGTGGAGAGATGTACTTCACGCGCGCGCGCGCGGAGGT 52798
QY 468 -----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGTCCACGGTCGATTGCGTACCGGTGGAGGCGGTTCACGAAGGCGCTGCGTGAGA 52858
QY 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLe 499
Db 52859 TCTGCTCCTGTCTGCAAGACACCCGCGCAGGTGTGGCAGAGGTTTCAGGCGGCTCCGCTGCC 52918
QY 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGCTCGAGCTTCTCTTTTCGCGCGCGGAGCCCGCGGAAGAGTGTGTACGCTTT 52978
QY 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCGAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGCTCTTCTCTCCGCG 53038
QY 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551
Db 53039 GGGGGCTTGTGCTGTGATGACACAGGCGGAGCAGAGGCGCTCGCTCGTATCGCTGCT 53098
QY 551 uIleGlyLeuAspArgGlyGluCys 560
Db 53099 GGAAGGG-----CGAGGCGAGGCGGTGC 53120
```

RESULT 11

```
US-10-152-886-44
; Sequence 44, Application US/10152886
; Publication No. US20030064491A1
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
; TITLE OF INVENTION: STRUCTURES
; FILE REFERENCE: 3011-3US
; CURRENT APPLICATION NUMBER: US/10/152,886
; CURRENT FILING DATE: 2002-05-21
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 5820
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
US-10-152-886-44
```

Alignment Scores:

```
Pred. No.: 0.00204 Length: 5820
Score: 123.00 Matches: 146
Percent Similarity: 31.73% Conservative: 59
Best Local Similarity: 22.60% Mismatches: 249
Query Match: 3.49% Indels: 192
DB: 13 Gaps: 32
```

US-10-691-383-2 (1-676) x US-10-152-886-44 (1-5820)

```
QY 4 HisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23
Db 1816 CAGCGCCAGCATCTACCGCGCGCGAACCTG---TCCACCGCGCGCGAGTGC 1872
QY 24 ThrSerGluGlnArgAlaLys-----GlyTrpArgArgGlnLeuGluGlyGluLys 40
Db 1873 ACCGACGTCGCCCGCGCGCATCGTCACCGGTTGCTGCGCGCGCTGCGGGTCTGAAG 1932
QY 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 1933 AGCTTCGGCATCGAGCGCGCCACGGTC-----ACGGCGCCACAGCTTCGCGAGCTCACC 1986
QY 61 -----TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeu 77
Db 1987 GCCTGCACCTGGCGCGCGCTCACCGAAGCGAGTGTCTCAAACTGGCCAGATCCG 2046
QY 78 GlyLysIleMetAlaArgVal----- 84
Db 2047 GCGAAGTGATGCGGACCGCGCAGCGGAGCGGGCCATGGCGGCGATCGCGGCCACG 2106
QY 85 ---ArgIleAlaThrAlaLeuAla----- 91
Db 2107 CCGAGTGTGCGCCAGGCGCTTGGCGGAGGCGGAGAGTGTCTATCGCGGGCTACAACGCC 2166
QY 92 -----ValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla----- 106
Db 2167 CCGAGCAGAGTGTCTCTTCCGACCGCGGAGCGGATCGACCGCGTGTGTCGCCGCTGCC 2226
QY 107 -----SerGly 108
Db 2227 CGCGCGGAGGGGTACCGCGCGCCGCATCAAGCTCTCGCACGCGCTTCCACTCGCCGCG 2286
QY 109 valPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln----- 124
Db 2287 GTCGTCCCGCGCCCGAGGCGCATGACCGGGGATCTGCCCGCATCGACTTCGCGCGGCTC 2346
QY 125 -----ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db 2347 GACCGCGCGCTGCTCTCCACGCTGACCGGTGAGTCTCTGCACCGCGCGGAGAGCTCGGC 2406
QY 141 -----ArgAspAsnValAlaPheAlaSerArg---ArgAspAlaAlaArgArgGlu 156
```

```
Db 2407 GATCTGCTGGCGACAGGTGGTCTCTCGCGCTCGCTTCGCTGAGCGCGCGGAAGTCC 2466
Qy 157 ArgAspGlyThrGlyThrValCysGlnIleThrAsnGlyGlu 170
Db 2467 GCCGAGCGCAGCGACCTGGTGTATCGAGTGGCGCGCGCGCGTGTCTACCGCGCTGCTC 2526
Qy 171 -----ThrAspLeuAlaThrMetPheHisIleSerLeuProHisAsp 184
Db 2527 GGCACCATCGCGCGCGTACCGGTGCTTTCGATCGACACCGACGCTGACGCTCGCG 2586
Qy 185 GluLeuGlyGlnValThrAlaAspAspPheAlaIle-----LeuGluAspCysIle 201
Db 2587 CCGTGTCTGAAGTTCGCGGTGCGCGTTCGCGTTCGCGCGCGCAGCTGGAGACGTCCACG 2646
Qy 202 Leu-----AsnGlyAspPheSerIle----- 208
Db 2647 CTGTTTCAGCGCGCGGTGCTGCGCGCGCTCGCGCGCGCGCGAGTTCGTTCTCTCGCC 2706
Qy 209 -----CysGluAspValPro-----AlaGlyAspProAlaGlyArgLeuValAsn 223
Db 2707 AGCCGTCGAGCGCGCGCGTCCATCGCGCGCTGCTGACCGCGACCGAGTGGCGGAG 2766
Qy 224 ProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIlePro 243
Db 2767 CCGCGGAGCGCGCGCGGAAAGCGCTCGGAAAGCGGTGGCAGCAGCACCC----- 2817
Qy 244 ProValProThrLeuSerSerProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMet 263
Db 2818 ---CTGACCTGTCTGCGC-----AAGCTCGCGTCCGAAAGCGGTGGCGCTCGAA 2868
Qy 264 AlaLeuAlaArgAspVal---ProPheMetGlnTyrGlyThrAspGluIleThr----- 280
Db 2869 GCGGTACCCGCGACACGATCCCTCGACGATCTGCACCTTTCGTGTCATCACGTCGCG 2928
Qy 281 -----ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsn 295
Db 2929 CAGCTGTCAACGAGTGACCGGCGCTCGCGCGCGCGCTGGAGGTATGCGGAAC 2988
Qy 296 LeuAspAlaValSerIleGlySer---AspGlyThrValAspProPheSerGlnLeuPhe 314
Db 2989 TTCGCGACCGTGTCTCGCGAACTCGCGGAGATGATGACGAGCTCGCGCGACCGCC 3048
Qy 315 Arg-----AlaThrPheValGlyValGluThrGlyProPheValSer 328
Db 3049 AACCCCGCGACACACAGCGCGAGTCCGCGCGTC-----GCCCGGTG----- 3096
Qy 329 GlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
Db 3097 -----GTCCGCGCGTTCGCGGTGGAGTACGTC---GTCCGCGCGAAGCG----- 3138
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 3139 ---TCGCGCGATCTCGCAGCGGTATCTCCACCGCGAGTGGACGCGTTCGCGCGCGCC 3195
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
Db 3196 GGTACACCA-----CTGCGGAGCGCGTGGCG-----GCG 3225
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
Db 3226 GCATTGSCCAGCGCGCGTTCGCGGACGGT----- 3255
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 3256 GTCTCTCTGCTGAAACGCGCAGCGGCTTCGCGGACGCTCGCGCGCTGTCTCGTGGCGCG 3315
Qy 429 AspArg-----GlnAlaGlyPhe 434
Db 3316 GCGCGCGGTGTCGCGCGCGCCCAACGCGCGGTTCGTGCTGTCGAGCAGGTCTC 3375
Qy 435 ValAsnPheGlyThrSerHisTyrPheArgLeu----- 445
```

```
Db 3376 GGCGCTCGCGGTGGCGAAGACGCTCCGCGTGGAGACCCGTCGCGCGCCACCGCATC 3435
Qy 446 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArg 465
Db 3436 GTGACCTCGCCGACCTCGGCCCG----- 3459
Qy 466 PheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
Db 3460 ---GTGACCCCGAAGCCCTCGACGCGCGGTG---TCCACTGTGGTCACCGAAGTGGCG 3513
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIle 505
Db 3514 GCGACGACCGACTTCAGCGAAGTCCGCTACGACACCGCGGAGTCCGCGCGTCCGGAAG 3573
Qy 506 AsnAlaAlaGlnAsnPro 511
Db 3574 CTCGCGCGCTGACCCCG 3591

RESULT 12
US-10-194-163-21/c
: Sequence 21, Application US/10194163
: Publication No. US20020172976A1
: GENERAL INFORMATION:
: APPLICANT: Ross, Bruce Carter
: TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
: AND USES THEREOF
: NUMBER OF SEQUENCES: 1120
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 PAGE MILL ROAD
: CITY: PALO ALTO
: STATE: CA
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSeq for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/194,163
: FILING DATE: 04-Nov-2002
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Basu, Shantanu
: REGISTRATION NUMBER: 43,318
: REFERENCE/DOCKET NUMBER: 529282000101
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-813-5995
: TELEFAX: 650-494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 21
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2843 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: UNKNOWN
: ORIGINAL SOURCE:
: ORGANISM: PORPHYROMONAS GINGIVALIS
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1...2843
: SEQUENCE DESCRIPTION: SEQ ID NO: 21
US-10-194-163-21

Alignment Scores:
Pred. No.: 0.00845 Length: 2843
Score: 122.00 Matches: 141
Percent Similarity: 35.52% Conservative: 113
Best Local Similarity: 19.72% Mismatches: 264
```

Query Match:	3.46%	Indels:	198
DB:	13	Gaps:	38
US-10-691-383-2 (1-676) x US-10-194-163-21 (1-2843)			
Qy 31 GlyTrrpArgArgGlnLeuGluGlyGluLysSerLeuGlyPheHisProSerGluThrPro 50			
Db 2190 GGTGGTGGCGACAAATGGAGATCGACGATATTGTCATCTTCGATCGTGGCGTGCCT 2131			
Qy 51 TyrIleLys-----TyrLeuGluGlySer 58			
Db 2130 CGGTGTACGCCGAACGTTTGACCGGGGGCTGGATCTGCTCCCTCTTTTGAAGAGCGGT 2071			
Qy 59 GluThrTrpIlybLysValLysLeu-----Pro 67			
Db 2070 CTGAGCTTCCTCTGCTCGTCTGTCAAAGCGCACTCGATGTATATCGTTGCGGTCCC 2011			
Qy 68 ThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg-----83			
Db 2010 GATACGGCTTCGAACATTTCAGTCTCTGCGATGCTCGAAAGATAAGAAATCGGAAGAG 1951			
Qy 84 --ValArgIleAlaThrAlaLeuAlaValValLeuAlaLaProCysLeuAlaPheAsp 102			
Db 1950 CCGAGCGGCTACGACTCAATATCAATACCATCTTGCTCAGGAACCTGC-----1903			
Qy 103 GluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHis 122			
Db 1902 CGGCTCTCTGCTTCGGGATTGTACAGACCCAAATC-CGACACATTTGGATTGCTCGA--1847			
Qy 123 LeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAsp 142			
Db 1846 ---CTTGAATGCCAAAGTACGTCT-----TTTCCAAGCGACAGTACGGGACTGAA 1799			
Qy 143 AsnValAla-----PheAlaSerArgArgAspAlaAlaArgGluAlaArgAsp 158			
Db 1798 AGCTCAGCTGGCCCATCTTTCTTTTCGAAAAAATACCGGCTCTCGATCTCAAAGA---1742			
Qy 159 GlyThrGlyThrValCysGln-IleThr---AsnGlyGluThrAspLeuAlaThrMetPh 177			
Db 1741 -----ACTGCATCCGATGTCAGTCTCGTGGCGACACTTTGGGAGTGGTCTCGTCT 1691			
Qy 177 eHisLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLe 197			
Db 1690 TCATATCTCTTTGGCGAATAGTCACATCGGTATGCTCTCCGA-----1648			
Qy 197 uGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspPr 217			
Db 1647 ---TACTGTGTTCTCTGCGCCGAAACTCTGCTTCGCCATAAGGTTGACGGACTT 1592			
Qy 217 oAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPh 237			
Db 1591 GGGCGGACGG---ATCCGTCCCTCC-----GACCTGAAGGTTTGCTCCC 1550			
Qy 237 eSerAlaThrThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLe 257			
Db 1549 GGCACTCGATACGATGGATGATATTAATGAATCTCTGTGCGCCAGCTGACAGTCGCTTC 1490			
Qy 257 uAlaGluLeuTyrrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrrGlyThrAs 277			
Db 1489 CGGCGAA-----GGGTTCTCGCTCGAAATCAGT---ATGAACATATGCG---GA 1448			
Qy 277 pGluIleThrThrThrAlaAlaAlaAsnLeuAlaGlyMetGly-----292			
Db 1447 TGAACCTGACCTGCAGCGCAGCATAGGGTTGGCAGAAATAGAAAAAGNACGGATGCA 1388			
Qy 292 -----292			
Db 1387 TGTCACTCCGGTTGGATGAGCTTAAGCTCGGAGTCCGGACCTACATGCTGGCAGAAGC 1328			
Qy 293 ----PhePro-----AsnLeuAspAlaValSerIle-----301			
Db 1327 TCTTTTCCCGATAAGCTCCCACTCAAATGCGCACCTTGGTGTCTCATGATATTATGCG 1268			

Qy 431 InAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluL 451
Db 2409 CGGTATCTGTCACAGC---CGTACGACAAACGGCGTGCACCGCTGTGGACCGACTTCAG 2465
Qy 451 euAlaGlnArgAlaSerCysTyrGlnLysTyrGlnValHisArgPheAlaArgPro 469
Db 2466 CGGCGCCCGGGCCACATGC---AGATCTGGGGCGGCGCCGAGTCGGGCAAGTCCA 2518

RESULT 14
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 428 Length: 9025608
Score: 122.00 Matches: 166
Percent Similarity: 31.57% Conservative: 73
Best Local Similarity: 21.93% Mismatches: 280
Query Match: 3.46% Indels: 240
DB: 15 Gaps: 39

US-10-691-383-2 (1-676) x US-10-156-761-1 (1-9025608)

Qy 4 HisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23
Db 498627 CACCTGGCCCGACCCACCGGGGGT---GTGACGACCTCGTGTGGGAAGCTGCTT 498683
Qy 24 ThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyGlySerLeuGly 43
Db 498684 CGGCTCTCCGAAGCGGTTCGGCGGATCACCGAGGCATACATCTCCGCGGGACCTC--- 498740
Qy 44 PheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThrTyrLysLys 63
Db 498741 -----ACCGAGCGCGGTGTCTCGCCCTATCTC-----GTAGAGAGCTTTCGTCAT 498785
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeu----- 77
Db 498786 GATTCTCTCCACATCTGCTCTCGGGTCTGGGTACTCGCGGTGGACGCGGGCGG 498845
Qy 78 -----GlyLysIleMetAlaArgVal-----ArgIleAlaThrAlaLeu--- 90
Db 498846 TCATGCGGTCTGTGCTCGCGAGTGTCTACGGGGCCGACGACATCGGTGACGGGTCCGC 498905
Qy 91 -----AlaValValLeuAlaAlaProCysLeu 99
Db 498906 CAGCCAGGAATCCACCGCGCGGTGTGTCGCGGTGGTCTCTCCCAA----- 498953

Qy 100 AlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHis----- 116
Db 498954 -----CATCGAGTAGTGGTTTCCGGTCACTTC---GACGATCGTGTCCACGCTGTCGGG 499004
Qy 117 -----ThrGlyGluGlyArgHis-----Leu 123
Db 499005 CGGTGCGCGCGCCGTCACGTCACCGGGCGCGCGCAGCATGGCTCGTCTCGCCCGTAC 499064
Qy 124 GlnThrCysThrAsnSerAspAlaLeuAspProThr-----Ala 137
Db 499065 GAGCAGGGTCCGCGCGCGCATCTCGCGCGCGCCAGTCTCGAACAGCGGCATATAGCG 499124
Qy 138 ProAsnArgArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArg 157
Db 499125 GCCCATCGCGTGAGTGTTCGCGCGCGCCTCATGAAGTCCAGCGCTCGAGTGGCG 499184
Qy 158 AspGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe 177
Db 499185 CTGGGCCATCGGTGAGCACCTGTGTGAAGGGGACCTTCGCGGTGAAGTGTGAC 499244
Qy 178 His-----LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp 193
Db 499245 CATCACACCGCGCGCGCTCCCAT----- 499274
Qy 194 PheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValPro 213
Db 499275 -----GGCTTCAGGTGTGCGCCACGCC--- 499298
Qy 214 AlaGlyAspProAla-----GlyArgLeuValAsnProThr 225
Db 499299 GTGGCGCATCCAGCCCGCGGAGTAGCCGAGCAGACGCGCGCTTCGCGCGGCACA 499358
Qy 226 AlaAlaPheAlaIleAspIle---SerGlyProAlaPheSerAlaThrThrIlePro--- 243
Db 499359 CTCCCGTACGCTGCGCGGTGCAGCGGAACACACGCTCGGGTTCGAGCGGACCGGCTC 499418
Qy 244 -----ProValProThrLeuSerSerProGluLeuAlaAlaG 256
Db 499419 GCCCTTCTCTAGCGGGGTGGACCGCGCCACACG-TCCAGCGGTTCGCGGAAGGGTG 499477
Qy 256 InLeuAlaGluLeuTyrTyrMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyT 276
Db 499478 CGGCGAACCGCGCGTACTGTGCGCGCCCGCGGTACGACG-----TACGCGC 499525
Qy 276 hrAspGluIleThrThr-----ThrAlaAlaAlaAsnLeuAlaGlyMetGly--- 291
Db 499526 AGAAGCACATCAGCGCGAGGGATGTGCTTCGCGACACCGCACCGGGCGCGCGCC 499585
Qy 292 -----GlyPheProAsnLeuAspAlaValSerIleG 302
Db 499586 GGTCCAGTTCTCGGGCGCTCGAACGACGCGCGGAAGCGCGCTCCATCAGCATCC 499645
Qy 302 LysSerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValG 322
Db 499646 CCATGGCTCGGGCGCGTCTCGCTTCGCTCGCGTCCGCTAGAGTGGACACGACATGT 499705
Qy 322 LuthrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrV 342
Db 499706 CGGC-GGGCGGTCTCTCCGCGCGCGCTCGCGTGTCT-----C 499743
Qy 342 alGluProLysGlnGluThrPheAla---ProAspLeuAsnTyrMetValAspPheAspG 361
Db 499744 TGGAGCGCGCTCAACAGTCTTGGCGAGCTTGGCGGCGTGGGATGTTCGAAGCCAGT 499803
Qy 361 LuthrLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGlu--- 379
Db 499804 CCGAGGCGGCGGCGACAGCGGTCCGCGCGCACAGTCCGCGCAGCAGCTCGATCGCGGTG 499863
Qy 380 -----GluLeuArgPheIleArgAsnAlaArgAsp---LeuAlaArgValS 394
Db 499864 AGCGAGTCGAAGCGAGCTCCAGGAAGTGGCGGTGGCGACGCTCTCTCGCGCGACGCG 499923

Qy 394 erPheValAspAsn-----IleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeu- 411
Db 499924 TGCCCGAGGACCGCGCTCTGGCGCGGACGCGTGTCCAGCAGCGCTCTCTCTGCTCG 499983
Qy 412 -----GluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheI 426
Db 499984 ACCGGGCGAGAGCGGCGAGTCGTCGCCAGCGCGCTCTCGCGGCT---GGTCCCTCG 500040
Qy 426 leAspSer-----AspArgGlnAlaGlyPheValAsnP 437
Db 500041 CGGACCGCGCTCTCGGCTCTCAGAGCGGCTGTACGTGGCGAGTCGCGCAGCAGCGG 500100
Qy 437 heGlyThrSerHis-----T 442
Db 500101 CTGGCGCTCCAGGTGAAGCGGCGGAGCACTTCTCCAGTCGATCGGCGGACGGTG 500160
Qy 442 yrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpG 462
Db 500161 AGGCGCGCTCTGCTGCTCGACGCGCTCGCGCAGCGCGGAGAT----- 500204
Qy 462 lnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaG 482
Db 500205 -----CGCAGCTCTCGGGGC-----CATGACGCCAGGCCCC 500235
Qy 482 lyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgV 502
Db 500236 CGCGCTCTCGCGC-----CTCTGAGCCGCG 500262
Qy 502 alAlaGluIleAsnAlaAlaGlnAsnPro-----AsnAsnGluV 515
Db 500263 TCGTCGAC---CATGCGCGCGCGCGCCAGCGCGCCAGCGAGGTCGCGGTCAAG 500319
Qy 515 alThrTyrLeuLeuProGlnAlaIleGlnVal-----G 526
Db 500320 CCGCGCGCGCGCGCGCGCGGTGTCAGAGGCGGTGGCGGAGGTAGGAG 500379
Qy 526 lySerProThr---HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheA 545
Db 500380 GCCTGCCACCGCTGCCCGCAGACCGCGCAGCGGCGGAGGAGGAGGCGTCCAGC 500439
Qy 545 laThrValLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProv 565
Db 500440 TCAGCGCGTCGAGCAACTCGTCGAGGTGGATGCGCGCGGTGAC----- 500483
Qy 565 alPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrG 585
Db 500484 -----CTTGAGCGCGCGCTTC----- 500501
Qy 585 luGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg-----GlnMetLeuG 603
Db 500502 --GGCGAAGCTCTCGAGACCGCGCTCGGC---CAGCATGAGACCGTGGCGCGCGCG 500556
Qy 603 lyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGlu 618
Db 500557 GCGTGCAGCAGCGCGGTGAGCGGTGCGGAGCTCGATCGCGCGGAG 500603

RESULT 15

US-10-437-963-65478/c
; Sequence 65478, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barzak, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 65478
; LENGTH: 5040
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_66522C.1
US-10-437-963-65478

Alignment Scores: 0.00286 Length: 5040
Pred. No.: 121.00 Matches: 159
Score: 30.40% Conservative: 93
Best Local Similarity: 19.18% Mismatches: 277
Query Match: 3.43% Indels: 301
DB: 17 Gaps: 39

US-10-691-383-2 (1-676) x US-10-437-963-65478 (1-5040)
Qy 26 GluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyLysSerLeuGlyPheHis 45
Db 3897 CAAGCAGGCTTGAGTACCTGAGCAACAAACAGAGAGTTACGATGACAGGGG----- 3844
Qy 46 ProSerGluThrProTyrIleLysTyrLeu-----GluGlySerGluThrTrpLysLys 63
Db 3843 -----CGGTAATTGATATTGTCATGGCAGCATGCTGATGTGTGG---AGG 3799
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle----- 80
Db 3798 GTTGCTGTTGATACCCAGGCGCTTGAGGCGCAACAAAATTTGTGGAACACTTGCAGATTTT 3739
Qy 81 -----MetAlaArgValArgIle----- 86
Db 3738 GTTCCATTTGCAAACTACAGACTGAGCGGAAGTTTGGGATTTTACCAATTAGATGCT 3679
Qy 86 ----- 86
Db 3678 TGCTCTTTTGTAGCAAAACATTTACGATGATGAAACCTTGTAGTATAGTAACATGATGC 3619
Qy 87 -----AlaThrAlaLeuAlaValLeuAlaAla-----ProCysLeuAlaPhe 101
Db 3618 TCTCCTCATGCTACCATGTTGCGGCGCATTTGCAGCTGCTTTTCATCCT----- 3571
Qy 102 AspGluValThrAlaSerGlyValPhePheProGluGluHisLysHisThrGlyGluGlyArg 121
Db 3570 GATGAACCTTGTCTCAATGAGGTTCACCG-----GGTGCA 3535
Qy 122 HisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArg 141
Db 3534 CAAGTGTATTCCTGCAAGATTGGCGATACCCGTTTAGGATCAATG----- 3490
Qy 142 AspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThrGly 161
Db 3489 -----GAGACAGGACAGGC 3475
Qy 162 ThrVal-----CysGlnIleThrAsn----- 168
Db 3474 CTGTTAGGCGCTTGTAGCTGCGAGTAGACACAAATGTCATCTGATTAACATGAGCTAT 3415
Qy 169 GlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln 188
Db 3414 GGAGAGCTACCTTCTCTCTGACTATGGCAGATTTATT-----GATCTTGTCTAGT 3364
Qy 189 ValThrAlaAspAspPheAlaIle-----LeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 3363 GAGGTGTGACACAGCATCGTATTATTATTATTAGTAGTCTCGAAACAAATGGTCTGCT 3304
Qy 207 SerIleCysGluAspValProAlaGlyAspProAla-----GlyArgLeu 221
Db 3303 TTAACACTGCTCGGTGCTCTGTTGGTACTAGTTCAAGCATTTATGGCGTTGGCGCTTAT 3244
Qy 222 ValAsnProThrAlaAla----- 227

Db 3243 GTCTCACCAGCTATGGCTGCTGGAGCTCACTGTGTGTCGTAAGCACCAGCTGAGGGGATG 3184
Qy 228 -----PheAlaIleAspIleSer 233
Db 3183 GAGTACACATGGTCTAGCCGAGGTCCTACTGCTGACGGGGATCTTGGTGTCTCCATTAGT 3124
Qy 234 GlyProAlaPheSerAlaThrIleProValProThr----- 247
Db 3123 GCACCG-----GGTGGAGCAGTGGCTCTCTGTACCAACATGGACACTTCAGTCTCGC 3073
Qy 248 -----LeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 3072 ATGCTCATGAATGAACCTTCCATGTCAFCACCTTCTGCTCGCGTGGGTGGTTCCTTCT 3013
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 3012 GTTAGTGCCATGAAGCTGAAGGCTGAAGGCTTCCATTAGCCCTTATATCTGTGAGGAAGGCAATT 2953
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMet-----GlyGly 292
Db 2952 GAGAAATACAGCTGCATCAATAAGTGTGCTGAGGAGAGCTAAACACAGGCGCATGG 2893
Qy 293 PheProAsnLeuAsp----- 297
Db 2892 CTTTGTGAAGTTGATAGGCTTTTGAATATGCTCAACAGGCTAAGGAATTGCCACTTGT 2833
Qy 298 -----AlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArg 315
Db 2832 TCTTATAGATATCATCATCATCAAGTGGC-----AAGCCAACTTCTAACTTAGAGGA 2779
Qy 316 AlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuValAsnSerPhe 335
Db 2778 ATATATCTACGTGGAGTAAACATGCGCGCAACAACTGAGTGGACTGTTCAA----- 2725
Qy 336 ThrIleAspAlaIleThrValGluProGlyGlnThrPheAlaProAspLeuAsnTyr 355
Db 2724 -----CTAGACCCCAAAATTTTCATGAGGATGCAAGTAACATCGAGCA 2683
Qy 356 MetValAspPheAspGluTrpLeuAsnIleGlnAsnGlyProProAlaGlyProGlu 375
Db 2682 TTGTTCCATTGAGGAGTGGCTGACGTGATCC-----ACTGATAGCTGTGTATC 2629
Qy 376 GluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPhe 395
Db 2628 AAAATTCCTGAGTACATAAATGGTGACAAACAATGGCGGCACCTTCAACATTGTTGTAAT 2569
Qy 396 ValAspAsnIleAsnThr-----GluAlaTyr----- 404
Db 2568 CTTGTTAATATCAGCAGCGGCTTTCATATTATGAAGTTTATGTTATAGATTGCAAAAGCA 2509
Qy 405 -----ArgGlySerLeuLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsn 422
Db 2508 CCATGGCGTGACCTATTTCGCGGTACCAATCACTGTATATAAGCCATTATGCTCTATCA 2449
Qy 423 Gly-----ProPheIleAspSerArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 2448 GGAGAGCCACCTGCATTGACACTTTCAAAT-----CTTTCCTTTAAGTCAGGT 2401
Qy 441 HisTyrPheArgLeu-----IleGlyAlaAla-----GluLeuAlaGln 453
Db 2400 CACATTGAAGAAGGTTTATAAATGTGCTATTGGAGCTTCATGGGTGGAAGTTACAAATG 2341
Qy 454 ArgAlaSerCysTyrGln----- 459
Db 2340 CGCATCAGCCTTTGATCTACTCCTAGGAGGTTCTTTTGGACACTGTTCAGATCTGTCCA 2281
Qy 460 -----LysTrpGln-----ValHisArg 465
Db 2280 CTGAAGCGACCTATTAAATGGGAGGCTGTGTCACTTCTCTTCCACCTTCTCTCAAGAT 2221
Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
|||:::|

Db 2220 TTCAGTTTTCTCTGTAGAAGGTGGCTTAACCTTTGGAGTTGTCCATAGCCAGTTCTGTGCT 2161
Qy 486 Ala-----AspPheAspIle----- 490
Db 2160 AGCGGAATTGCTAGCCCATGAACCTACCTGTGTGTTGATTGAGATTGTGTTTCATGGAATA 2101
Qy 490 ----- 490
Db 2100 TCCGTTGATCAGAAAAATAATTGGCCTTGTATGGTAGTGAGGCACCTGTGCGTGTGTTGCT 2041
Qy 491 ---SerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaGln 509
Db 2040 AGATCTCTATTAGCATCAGAGAGCTTGTCTCT---GTTGCCACTCTGAAACAAGTCAAG 1984
Qy 510 AsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 529
Db 1983 ACCCTTATCGACAGTGGAGTCTAACTTTTGTTCACCTT-----CCTCCCAGC 1936
Qy 530 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 549
Db 1935 CGTGATAGATTACCATCTGGC-----AAGCAAAATCATAGCACTAACTCTGACT 1888
Qy 550 AlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro-----Asn 563
Db 1887 TACAAATTCAAACTAGAGGATGGTGTGAAATAAAGCCCGCTGTGCCCTACTTAACAAC 1828
Qy 564 ProValPheProSerAspGlyLeuGlu-----LeuIleAsnPheGluGly 579
Db 1827 AGAATATAT-----GATAACAAGTTCGAGTCTCAGTATTATAGAAATATCAGATTCAAAC 1774
Qy 580 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599
Db 1773 AAGTGTGCTACTCTTCTGCTGACGTCTATCTAATTTATGTCAAACTCTCAAAAGGTGAA 1714
Qy 600 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619
Db 1713 TACACGTGCAATTGTATACATAAGCATGACAATGTGCAG-----TTATT-GGAGAAACT 1661
Qy 620 IleThrValArgThrLeuHisGlnGlu 628
Db 1660 GNAGCACTTGGTCTATTATTATCGAGAG 1634

Search completed: September 19, 2004, 10:43:49
Job time : 4388 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 04:00:38 ; Search time 145.41 Seconds
(without alignments)
2579.931 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCXAADTRGSPWPTGVL.....SIDGCMGLVYTGVAQCQA 676

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPO.spool/US10691383/runat 17092004 102657 1766/app query.fasta_1.1230
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691383 @CGN 1 1 93 @runat 17092004 102657 1766 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3528	100.0	2931	3	US-09-151-189-1
2	3528	100.0	2931	4	US-09-596-794-1
3	129.5	3.7	11679	4	US-09-328-352-1377
4	123.5	3.5	68750	3	US-09-335-409-1
5	123.5	3.5	68750	4	US-09-568-102-1
6	123.5	3.5	68750	4	US-09-567-969-1
7	123.5	3.5	68750	4	US-09-568-480-1
8	123.5	3.5	68750	4	US-09-568-486-1
9	123.5	3.5	68750	4	US-09-568-472-1
10	123.5	3.5	68750	4	US-09-567-899-1
11	122	3.5	2843	4	US-09-221-017B-21
12	117.5	3.3	4411529	3	US-09-103-840A-1

13	117	3.3	2208	4	US-09-252-991A-15601	Sequence 15601, A
14	117	3.3	3132	4	US-09-252-991A-15639	Sequence 15639, A
c 15	117	3.3	53500	4	US-09-266-965-76	Sequence 76, Appl
c 16	116.5	3.3	1830121	4	US-09-557-884-1	Sequence 1, Appli
c 17	116.5	3.3	1830121	4	US-09-643-990A-1	Sequence 1, Appli
18	114	3.2	3270	4	US-09-252-991A-3763	Sequence 3763, Ap
19	113.5	3.2	10419	4	US-09-408-020-3	Sequence 3, Appli
20	113.5	3.2	42432	4	US-09-408-020-2	Sequence 2, Appli
c 21	112.5	3.2	1791	4	US-09-252-991A-15622	Sequence 15622, A
c 22	111.5	3.2	1527	4	US-09-252-991A-15698	Sequence 15698, A
c 23	110.5	3.1	1848	4	US-09-252-991A-11564	Sequence 11564, A
c 24	110	3.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 25	110	3.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 26	109.5	3.1	71989	4	US-09-443-501A-2	Sequence 2, Appli
27	108.5	3.1	6948	4	US-09-543-681A-1262	Sequence 1262, Ap
28	108	3.1	25165	4	US-09-453-702B-39	Sequence 39, Appl
29	107.5	3.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 30	107	3.0	1629	4	US-09-252-991A-3705	Sequence 3705, Ap
31	106.5	3.0	1308	4	US-09-252-991A-13571	Sequence 13571, A
c 32	106.5	3.0	4131	4	US-09-252-991A-13773	Sequence 13773, A
33	106.5	3.0	8211	4	US-09-252-991A-13656	Sequence 13656, A
34	106	3.0	4647	4	US-09-252-991A-5730	Sequence 5730, Ap
c 35	106	3.0	9515	1	US-08-920-812-13	Sequence 13, Appl
c 36	106	3.0	9515	1	US-08-920-827-13	Sequence 13, Appl
c 37	106	3.0	9515	1	US-08-921-177-13	Sequence 13, Appl
c 38	106	3.0	9515	1	US-08-362-577C-13	Sequence 13, Appl
c 39	106	3.0	9515	2	US-08-920-828-13	Sequence 13, Appl
c 40	105.5	3.0	3492	4	US-09-489-039A-3892	Sequence 3892, Ap
41	105.5	3.0	10480	4	US-09-732-615-13	Sequence 13, Appl
42	104.5	3.0	1158	4	US-09-252-991A-8509	Sequence 8509, Ap
43	104.5	3.0	1912	3	US-08-868-435-11	Sequence 11, Appl
44	104.5	3.0	1912	4	US-08-744-231-11	Sequence 11, Appl
c 45	104.5	3.0	2019	4	US-09-252-991A-8334	Sequence 8334, Ap

ALIGNMENTS

RESULT 1

US-09-151-189-1
; Sequence 1: Application US/09151189
; Patent No. 6233457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-09-151-189-1

Alignment Scores:			
Pred. No.:	0	Length:	2931
Score:	3528.00	Matches:	676
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-691-383-2 (1-676) x US-09-151-189-1 (1-2931)

Qy 1 MetLeuCysHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20
|||||

Db 228 ATGCTTTGCCATGCGCGGACACGACAAAGAGCTCTCTCTATGCTGACACCGGAGTGCTT 287
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaValGlyTyrArgArgGlnLeuGluGlyGluLys 40
Db 288 CGGTGTCTCATCAGACGAGCGCGCTTAAGTTGGAGACGCCAGTTAGAGGGGAGAAA 347
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrTyrIleLysTyrLeuGluGlySerGluThr 60
Db 348 TCACTAGTTTTCATCAAGCGAGAGCGCTTATATCAAGTACTTGGAGGCTCTGAGACT 407
Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
Db 408 TGGAGAAGGTTAAGCTTCCAAACGACGCGATATCGGCTTCCAAAGATCCTGGGTAAAT 467
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db 468 ATGCCACGGGTCCGATCGCTACCGCTTGGCTGTGTACTTGGCCGACCCCTGTTTGGCA 527
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
Db 528 TTCGACGAGTTCAGCCAGTGGTGTTCCTTGAGGAAACACAGCACACCGGGGAGGA 587
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db 588 AGACACCTCCAGACCTGTACAACTCCGACGATCGCTGGATCCGACGCGCGCGAATAGA 647
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAlaAlaArgArgGluArgAspGlyThr 160
Db 648 AGGACAACTAGTCTTTCGCTCGCGCGGATGCCCGGAGGAGAACGTGACGGGACA 707
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
Db 708 GGGACTCTGTCCAAATCACTAACGGAGAACTGATTTGGTACCATGTTCCCAAGTCT 767
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
Db 768 CTGCCACACGATGAATGGGACAGGTAAACGACGACTTCGCTATCTCCGAGGACTGC 827
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
Db 828 ATCTTAAACGGAGATTTTCAGATTTTGGAGAGCGTCCGCGGAGACCCCGCGGTCGC 887
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 888 CTGCTCAATCTACCGCTGGCTTGGCATCGACATATCCGGTCCCGCATCTCGGCTACG 947
Qy 241 ThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 948 ACAATACCCCGGTACCTTACCTTCTCTGAGCTCGCGCTCAGTTTGGCGGAGCTA 1007
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 1008 TACTGGATGGCGCTGGCCAGGAGTACCTTTATGAGATATGGCACCGACGAATATACC 1067
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
Db 1068 ACTACCGCGGAGCGCAACCTCGCTGGATGGAGGCTTCCAAATCTGGACGCGTCTCG 1127
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
Db 1128 ATAGGTCGAGTGTACGTTGGATCGGATCTCCAGCTCTTCCGAGCGACCTTCGTTGGT 1187
Qy 321 ValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle 340
Db 1188 GTTGAACCGGGCCCTTGTCTCTCAGCTGTCTGTGAACAGCTTCAACATCGACGCTATT 1247
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
Db 1248 ACGTTCGACCGAGCAGGAGACATTCGCCCGCCGACTTGAATATATGTCGATTTTAC 1307
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
Db 1308 GAATGGCTGAACATTCAGATGGTGGACCCCGCGCGCGCCGAGAGTTAGACGAAGAG 1367

Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
Db 1368 CTGCGTTTATCGTTAAACCGCCGACCTGGCCAGGCTCTCTCTCGTGAGCAATATCAAC 1427
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
Db 1428 ACCGAAGCTTATCGCGGCTCTCTTATCTACTTGAAGCTGGGAGCTTCAGACGCGCGT 1487
Qy 421 IleAsnGlyProPheIleAspSerArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 1488 ATCAACGGTTCATTCAGACGATCGCAGCGGCTTCGTCACTTCGACGCTCT 1547
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
Db 1548 CACTACTTCAGATTGATAGTGCCTGCGCAGCTGCGCTGCTGTATTACCAAAAG 1607
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
Db 1608 TGGCAGGTGCATCATTTGTCAGCCCGGAGGCTTCGCGGCTACCTCCCAACACCATC 1667
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
Db 1668 GCGGGGATCTAGATGACACTTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAAA 1727
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
Db 1728 CGTGTGGCGAGATTAATGCGGCGAGATCCCAACAGAGGTCACTTACCTTCTTCCA 1787
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 1788 CAAGCTATCCAACTGGGATCGCCAAAGCACCCTTCTTACCCGTCGCGCACGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
Db 1848 AATGGAGCATTTGCCAGATTCTGAAGGCCCTCATTTGGCCCTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 580
Db 1908 TTCCCTTAAACCCCGTGTCCCAAGGATGACGGCTTGAACATACTCACTTCGAAGGGGCA 1967
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCCCTTACATATCAGGAGAGATCAACAGCTCGCGTCAACGTCGATTTGGAGGAGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGATCCACTATCGTTTCGACGGTATCAAGGCCCTACTTCTCGAGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
Db 2088 ACTGTACGAACACTTCACGAGGAGCTGATGACGTTGCGCGGAGGAGCCACTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTGATCAAACTTTTCCAGGACGGGACATTCCTCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGCGGACTGCCAGGCT 2255

RESULT 2

US-09-596-794-1

; Sequence 1, Application US/09596794

; Patent No. 6656715

; GENERAL INFORMATION:

; APPLICANT: Vreeland, Valerie

; APPLICANT: The Regents of the University of California

; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases

; TITLE OF INVENTION: and Their Uses

; FILE REFERENCE: 023070-087110US

; CURRENT APPLICATION NUMBER: US/09/596,794

; CURRENT FILING DATE: 2000-06-19

;; PRIOR APPLICATION NUMBER: US 09/151,189
;; PRIOR FILING DATE: 1998-09-10
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 2931
;; TYPE: DNA
;; ORGANISM: Fucus distichus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (228)..(2258)
;; OTHER INFORMATION: vanadium bromoperoxidase
US-09-596-794-1

Alignment Scores:

Pred. No.: 0 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-691-383-2 (1-676) x US-09-596-794-1 (1-2931)

Qy	1	MetLeuCyHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu	20
Db	228	ATGCTTTGCCATGCAGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTGCTT	287
Qy	21	ArgLeuLeuThrSerGluGlnArgAlaValGlyTTPArgArgGlnLeuGluGlyLys	40
Db	288	CGGTGTCTCATCAGACGACGCGCTTAAAGTTGGAGACGCCAGTTAGAGGGGGGAGAA	347
Qy	41	SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr	60
Db	348	TCACTAGTTTTCATCAACGGAGACGCTTATATCAAGTACTTGGAGGCTCTGACACT	407
Qy	61	TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle	80
Db	408	TGGAAGAAGGTAAAGCTTCAACCGACGCGCATATCGGCTTCCAGATCCTGGGTAAAT	467
Qy	81	MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla	100
Db	468	ATGGCCAGGGTCCGCACTCGCTACCGCTTGCTGTGTACTTGGCCGACCCCTGTTGGCA	527
Qy	101	PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly	120
Db	528	TTCCAGCAGGTTCAGCCAGTGGTGTTCCTTGAGGACACACAGGACACACCGGGGAGGA	587
Qy	121	ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg	140
Db	588	AGACACCTCCAGACCTGTACAAACTCCGACGATCGCTGGATCCGACGGCGCGCAATAGA	647
Qy	141	ArgAspLeuValAlaPheAlaSerArgAspAlaAlaArgGluArgAspGlyThr	160
Db	648	AGGACAACTGAGCTTTTGGCTCGCGCGGATGCCCGCAGGCGAGAACGTGACGGGACA	707
Qy	161	GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer	180
Db	708	GGGACTCTCTGCCAAATCACTACGGAGAACTGATTTGGCTACCATGTTCCACAGTCT	767
Qy	181	LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys	200
Db	768	CTGCCACACGATGAATGGGACAGGTAAACCGACAGCTTCGCTATCCTCGAGGACTGC	827
Qy	201	IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg	220
Db	828	ATCTTAAACCGGAGATTTTCAGCATTTGCGAGACGTGCTCGGGAGACCCCGCGGGTCG	887
Qy	221	LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr	240
Db	888	CTGCTCAATCTACTCCGCTGGTTCATCATGACATATCCGGTCCCGCATCTCGGCTACG	947
Qy	241	ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu	260

Db	948	ACAAATACCCCGGTACCTACCTTTCTCTAGCTCGCGCTCACTAGTTGGCGAGGTA	1007
Qy	261	TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr	280
Db	1008	TACTGGATGGCGTGGCCAGGATGTACCTTTATGAGATATGCGACCGAGCAATATACC	1067
Qy	281	ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer	300
Db	1068	ACTACCGCGGACGCAACCTCGCTGGAATGGAGGCTTCCCAATCTGGACGCGGTGCG	1127
Qy	301	IleGlySerAspGlyThrValAspPropheSerGlnLeuPheArgAlaThrPheValGly	320
Db	1128	ATAGGGTCGATGGTACGGTGGACCCGCTTCTCCAGCTTCTCCAGGACCTTCTGTGT	1187
Qy	321	ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle	340
Db	1188	GTTGAAACCGGGCCCTTTGTCTCTCAGCTGCTGCTGAAACAGCTTCACTACGCGCTATT	1247
Qy	341	ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp	360
Db	1248	ACGGTCGAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGGTCGATTTTGAC	1307
Qy	361	GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu	380
Db	1308	GAATGGCTGAACATTCAGAAATGGTGGACCCCGCGCGGCCCGGAGAGTTAGACGAAGAG	1367
Qy	381	LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn	400
Db	1368	CTGGCTTTATCCGTAAACCGCGGACCTGGCCAGGGTCTCTCTTGTGGACATATCAAC	1427
Qy	401	ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly	420
Db	1428	ACCGAGCTTATCGCGGCTCTCTTATCTACTTCTGAGCTGGGAGCTTTCAGCAGGCCGCT	1487
Qy	421	IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer	440
Db	1488	ATCAACGGCTCATTCATCGACAGTATCGGACGCGGGCTTCGTCAACTTCGGCACGCTCT	1547
Qy	441	HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys	460
Db	1548	CATCTTCTAGATGTAGGTCCGCCAGCTGGCGAGGTGCTGCTGTGTACAAAG	1607
Qy	461	TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle	480
Db	1608	TGCGAGGTGCATCGATTTGACGCCCGAGGCTCTCGGGGTTACCCCTCCACACACCATC	1667
Qy	481	AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys	500
Db	1668	CGCGGGGATCTAGATGTCAGACTTCGACATCTCCCTTCTTGAAAATGATGAGCTCTTGA	1727
Qy	501	ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro	520
Db	1728	CGTGTGGCGGAGATAATGCGGCGAGATCCCAACAAACGAGGTCACTACTCTTCTTCCA	1787
Qy	521	GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln	540
Db	1788	CAAGCTATCCAGTGGGATCGCAACGACCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1847
Qy	541	AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys	560
Db	1848	AATGGAGCATTTGCCACAGTCTTGAAGGCCCTCATTTGGCCCTAGATCGGGGAGGTGAGTGC	1907
Qy	561	PheProAsnProValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAla	580
Db	1908	TTCTCTAACCCGCTTCTCCCAAGGATGAGCGCTCGAATCAATCACTTCCAAAGGGGCA	1967
Qy	581	CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln	600
Db	1968	TGCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGCTCGCATTTGGGAGGCGAG	2027
Qy	601	MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle	620


```
Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
Db 2215 -----AGCGTACCAGGTAGTACTTAAC 2238
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuysArgValAlaGluIle 505
Db 2339 GCGATGCGAGT-----AAGACGATTGATGCTAAAGTAAACGTTTACA 2280
Qy 506 AsnAlaAlaGlnAsnPro-----AsnAsnGluValThrTyrLeuLeu----- 519
Db 2281 GATCGCGCAGGTAATAGCAGCAGGTGTTAAACGATACACAAATATACATCGATACCAC 2340
Qy 520 ---ProGlnAla-----Ile 523
Db 2341 GCACCTGATGCACCAAGTAATTAACCGGTTAAACGGGACAGACCGCGATTACAGGTACGGCA 2400
Qy 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
Db 2401 GAGCTGTTCAACAGTAACGTGCTACTTATCCAGATGCGAGTACAA----- 2448
Qy 544 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsn 563
Db 2449 -----ACAGTTGTTGACGACCGGAT---GGCAGTTGCAGACAGTACCAAC 2490
Qy 564 ProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThr 583
Db 2491 CCAGTTTAAATGATGCGGAT----- 2511
Qy 584 TyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGly 603
Db 2512 -----AAAGTTACAGCAATTGCTACAGATCCAGCA---GGCAACCCATCATTAACCA 2559
Qy 604 IleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArg 623
Db 2560 GGTACACACTAGTGTGATGAGTGGTGGTCCAAATACCGATGGTGTATTTACCGTTGAT 2619
Qy 624 ThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArgLeuPhe 643
Db 2620 TCAGTAACAGCTGCAATATGATTAATGCTATCAAGACGATCAGGCAACGTTACTGT--- 2676
Qy 644 ThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCys 663
Db 2677 ACAGGTGATTGAAAAACGTTCCGGCAGATGCGCAAAATACAGTGTGCTGTTGTGATC 2736
Qy 664 SerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2737 AATGCCACAGCTATACTGCAACTGTATAGCACAGCA 2775
```

RESULT 4

```
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt. Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

Alignment Scores:

Pred. No.: 0.704 Length: 68750

```
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 3 Gaps: 26

US-10-691-383-2 (1-676) x US-09-335-409-1 (1-68750)

Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGCTGTGGAGCGCGCGCTGGAGCGCGCGAGCGCGCGCGGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGTTCGAC-----CTGGTGTCTGGCAGCTTCCATGCGAAGTGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGAGAGCGCTCGCGCAGGCGGTACATCTCGGCACTCTCGCATATGGAAGCTTCTG 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrC 126
Db 51681 GCT-----GCTGGAGAGCGCTCACATAGACGAGTTG 51713
Qy 126 sThrAsnSerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaP 146
Db 51714 CTGTCAGGCTTCAATCTCTGCTCTACAGGAAGGTCATCAAGCGATGATGGAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
Db 51774 CTTGTGCGCA-----TCGGCATCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCCGTCGCCGAGCGCTGATTTGGCGGCTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGAGGGTGTTCGCCGACCTCCAGTCTCTATTATTAGTGTGTCAGTGTTCGCGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGCGAGCTATTGACCGGTAAAGCGCTCGGCTCGAGATCTCTTCCCTGTGTGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTCGCCCATCGCCGTTACTCGAACGGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
Db 52049 CGTGGCGGTGTCGTTCGAGTCGCGCGCGGTGGTAGCACCGTCGGGAATGTTTCAGCAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGGAGATCGAGACGAGGACGCGCGCGACCGCGCGCTCTCCCGGTGTGCTGCC 52168
Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGACGAGTACCATTTCCACGATGTTTCCGCTCTTCTCTCGCGGGAGCA 52228
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAl 284
Db 52229 AAGATTTGAGATTTATTCATTCCTGAAGTATGTCATTCGTGATTCGACCGAGGAGCAGC 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyPhePheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGGATACGCACATCAGAGGTTTGACGTCATCTCGCGGCCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrG 324
Db 52346 TGGACCCCGCAT-----ATAAGGCCAGC----- 52370
Qy 324 yProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
```

```

; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-568-102-1 (1-68750)

Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
|||:|||||
Db 51516 AGTCACACCTGGCTGTGGGACGGCGCTGGACGGTGGACGGCCGACGGCGGAGCGTG 51575
|||:|||||

Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
|||:|||||
Db 51576 CCCTGTCAC-----CTGGTCTGGCAGCTTCCATGCGAAGTGGAGGTC 51620
|||:|||||

Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
|||:|||||
Db 51621 ATGAGGCGCTCGCGCAGGCGTACATCATCGGCACCTCTCGCATATGAAACGTTCTCTGC 51680
|||:|||||

Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCys 126
|||:|||||
Db 51681 GCT-----GCTGGAGAGCGTCCACGATAGACGAGTTG 51713
|||:|||||

Qy 126 eThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146
|||:|||||
Db 51714 CTGTCAGGCTTCAATCTCTGCTCTACAGGAGTCTCAAGCATGGATGGACAC 51773
|||:|||||

Qy 146 eAlaSerArgArgAspAlaAlaAlaArgArgGluArgGlyThrGlyThrValCysGlnI 166
|||:|||||
Db 51774 CTGTGCGCA-----TCGGCATCTT-GTAGGGACGGAGAGCATTTGTGAGCTCTCA 51826
|||:|||||

Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
|||:|||||
Db 51827 GCCGCTGCCGAGCGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
|||:|||||

Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
|||:|||||
Db 51869 CGGAGGGTGTTCGCCGACCTCCAGTCCTATTGTAGTGTGCAAGTTTCCGGGGAAACG 51928
|||:|||||

Qy 201 -----IleLeu-----AsnGlyAs 205
|||:|||||
Db 51929 GCTCGCGGACGTATTGACCGGTAAAGACGCTCGCGCTCGAGATCTCTTCTCGTGTGCTC 51988
|||:|||||

Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
|||:|||||
Db 51989 GTTCGATATGCGGAGCGAATCTATCGAGATTGCCCATCGCCGTTACTCGAACGGCAT 52048
|||:|||||

Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
|||:|||||
Db 52049 CGTGGCGGTGTGCTGAGTCGCGCGCGGGTGTAGCACCCTCGGGAATGTTTCAGCAT 52108
|||:|||||

Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
|||:|||||
Db 52109 CTTGAGATCGGAGCGGAGCGGCGGACACCGCGCGCTCTCCCGGTGTGTCTGCC 52168
|||:|||||

Qy 250 rProGluLeuAlaAlaGlnLeuAlaGluLeuValTrpMetAlaLeuAla----- 266
|||:|||||
Db 52169 TGACCGGAGGAGTACCATTTCAACGATGTTTCTCCGCTCTTCTCTGTCGCGGAGCA 52228
|||:|||||

```

```

; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goslach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```

```
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284
Db 52229 AAGATTTTCAGATATTCATTCCTGAAGTAGTGGCATTCCTGGATGTCAGCAGGAGCCAGC 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGGATACGCACATCAGAGGTTTGAGCGTATCTGTCGGCGGCCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPhePheAlaThrPheValGlyValGluThrG1 324
Db 52346 TGGCAGCCCGCAT-----ATAAGAGCCAG----- 52370
Qy 324 yProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGAAGCGTCTCCTG----- 52385
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTGTCTCGCGCCGAGGCCCTTCGTGCTGTCGAGGGGCAAGGGCATCC 52438
Qy 360 pGluTyrLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluG1 380
Db 52439 GATCTGCTTCATATCACCG---GGATTGATGAGGGTGGCAGAGTACGAGATGA 52495
Qy 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCTCTGCTCGACCTGGTGTGACGCTCTCGCGCCGGT 52555
Qy 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGCTTTTCGCGAGCCGCTGAGTCTGCGACGCGAGTCTCCGCGGGGATCTCCGACA 52615
Qy 407 xLeuIleLeuLeuGluGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427
Db 52616 GCAGTGATCTCTCG-----CGGCGCCGGGATAGCAGAGCCGCTTGTGA 52663
Qy 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleG1 447
Db 52664 CAGCTCCGGTGAGTCG----- 52679
Qy 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTyrGln----- 462
Db 52680 -GCGACCGAATTCGCGCCGCGCGCTGCAGTACGACGAGGAATGGCGCATGCTCCGCTGA 52738
Qy 463 ---ValHisArgPheAla----- 467
Db 52739 CGTCGTCCATCGGATGCGTTGGAGAGATGTACTCCAGCGCGCGCGCGCGCAGGT 52798
Qy 468 -----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGGTCCACGGTCGATTGCTACCGTGGAGGCGCTTCACGAAGGCGCTCGCTGGAGA 52858
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLe 499
Db 52859 TCTGCTCTCTGTTCAAGACACCCGCGGAGGTCGTGGCAGAGGTTTCAGGGGCTCCGCTGCC 52918
Qy 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGTCAGGCTTCTCTTCGCGCGCGGAGCCCGCGGAGAGGTTGTGTACGCTTT 52978
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCGCAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGTCTTCTTCTCCGCG 53038
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551
Db 53039 GGGGCGCTTGGCTCGTGTGATGGACGAGCGGAGCGGCTCGCTCGTGTGCTGCT 53098
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560
Db 53099 GGAAGGG-----CGAGCGGAGCGGTGC 53120
```

RESULT 6

```
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1

Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-567-969-1 (1-68750)
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGTGTGGGAGCGCGCTGGAGCGGCCGAGCGCGAGCGGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGTTCGAC-----CTGGTCTCGGAGCTTCCATGCGAAGTGGGAGGTC 51620
Qy 87 AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGAGCGCCTCGCGAGCGGTACATCATCGGCACCTCTCGCATATGGAACGCTCTCTGC 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGlyArgHisLeuGln-ThrCy 126
Db 51681 GCT-----GCTGGAGAGCGTACACGATAGACGAGTTG 51713
Qy 126 sThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAspAsnValAlaPh 146
Db 51714 CTGTCAGGCTTCAATCTCTGCTGTACAGGAGGTTCATCAAGCATGGATGGAACAC 51773
Qy 146 eAlaSerArgArgAspAlaAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
Db 51774 CTTGTCCGCA-----TCGGCATCCTT-GTAGGGGACGAGAGCATTTTGTGAGCTCTCA 51826
Qy 166 eThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCCGTCGCCGCGCTGATTGTCGGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys----- 200
Db 51869 CGGAGGAGGTTTCGCCGACCTCCAGTCCTATTGTAGTGGTGAAGTTTCCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGCGAGCATTTGACCGGTAAAGCGCTCGCGCTCGAGATCTCTCTTCCCTGCTGGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTTCGATATGCGGAGCGGAATCTATCGAGATTGCCCATCGCCGCTTACTCGAACCGCAT 52048
```

Qy	218	-----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI	230
Db	52049	CGTGGCGGTGTCGTGAGTCGGCGCGGGTGTAGCACCGTCGGGAAGTTTCAGCAT	52108
Qy	230	eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe	250
Db	52109	CTTGGAGATCGGAGCAGGACGGGCGCGACCAACCGCGCGTCTCTCCGGGTGTGCTGCC	52168
Qy	250	rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla	266
Db	52169	TGACCGGACGGAGTACATTTCACCGATGTTTCTCGCTCTTCTTGTCTCGCGGGAGCA	52228
Qy	267	-----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAl	284
Db	52229	AAGATTTCGAGATTATTCATTCTCTGAAGTATGGCATTCCTGGATGTCCGACGAGGCCAGC	52288
Qy	284	aAlaAsnLeuAlaGlyMetGlyClyPheProAsnLeuAspAlaValSerIleGlySerAs	304
Db	52289	TGGCCAGGGATACGCACATCAGAGGTTTTCAGCTCATCTCGTCGGGCCAATGTC---ATCCA	52345
Qy	304	pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGl	324
Db	52346	TGGCACCCGCGAT-----ATAGAGCCACG-----	52370
Qy	324	yProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPr	344
Db	52371	-----GCCAAGCGTCTCTG-----	52385
Qy	344	oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs	360
Db	52386	-----TCGTTGCTCGCGCCGAGGCGCTCTCGTGTGTGTGAGGCGCACAGGGCATCC	52438
Qy	360	pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGl	380
Db	52439	GATCTGGTTCGATATCACCACG---GGATTGATTGAGGGGTGCACAGAGTACGAAGATGA	52495
Qy	380	uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa	393
Db	52496	TCTTCGTATCGACCATCCGCTCTGCTCTCGGACCTGTGTGTGACGTCCTCGCGCGGGT	52555
Qy	393	lSerPheValAspAsnIle-----AenThrGluAlaTyrArgGlySe	407
Db	52556	AGGTTTGGCGAGCCGCTGAGTGTCCAGCGCAGGATCTCGGCGGGGATCTCCGACA	52615
Qy	407	rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs	427
Db	52616	GCAGTGATCTCTCG-----CGCGCGCGGCATAGCAGGAGCGGCTTGTA	52663
Qy	427	pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGl	447
Db	52664	CAGCTCCGCTGAGTCG-----	52679
Qy	447	yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlybTrpGln-----	462
Db	52680	-GGACCGCAATCCCGGCCCGCGTGCAGTACGGCAGGNAUGGCGCGATGGCTCCGCTGA	52738
Qy	463	----ValHisArgPheAla-----	467
Db	52739	CGTCGTCCATCGGATGGCTGTGGAGAGGATGTACTTCCACCGCCGCGCGCGCAGGT	52798
Qy	468	-----ArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAs	483
Db	52799	TTGGGTCACGGTTCGATTGCGTACCAGGTCGAGCGCGGTTTCACGAAGGCGCTCTCGCTGGAGA	52858
Qy	483	pLeu-----AspAlaAspPheAspIleSerLeuLeuAsnAspGluLeuLe	499
Db	52859	TCTGCTCTTTCGAAGACACCGGGCAGGTCTGTGCAGAGGTTTCAGGGGCTCCGCTGCC	52918
Qy	499	uLysArgValAlaGluIleAsnAlaGlnAsnProAsnAsnGluValThrTyrLeuLe	519
Db	52919	GCAGCTCGAGGCTTCTGCTTTTCGCGCGCGGACCGCGGGAAGAGTGGTGTACGCTTT	52978

Qy	519	uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaLath	539
Db	52979	GGAATGGCAGCGCAAAAGACCCCTATACACAGAGGCTCCGACAGCGCGTCTTCTCTCGC	53038
Qy	539	r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe	551
Db	53039	GGGGGCTTGGCTCGTCTCTGATGGACACAGGCGGACAGCGCTCGTCTGATCGTCTCT	53098
Qy	551	uileGlyLeuAspArgGlyGluCys	560
Db	53099	GGAAGGG-----CGAGGCGAGGCGTGC	53120
RESULT 7			
US-09-568-480-1			
; Sequence 1, Application US/09568480			
; Patent No. 6355458			
; GENERAL INFORMATION:			
; APPLICANT: Schupp, Thomas			
; APPLICANT: Ligon, James			
; APPLICANT: Molnar, Istvan			
; APPLICANT: Zirkle, Ross			
; APPLICANT: Cyr, Devon			
; APPLICANT: Goerlach, Joern			
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
; FILE REFERENCES: 4-30582A			
; CURRENT APPLICATION NUMBER: US/09/568,480			
; CURRENT FILING DATE: 2000-05-10			
; PRIOR APPLICATION NUMBER: 09/335,409			
; PRIOR FILING DATE: 1999-06-17			
; NUMBER OF SEQ ID NOS: 30			
; SOFTWARE: Patentin Ver. 2.0			
; SEQ ID NO 1			
; LENGTH: 68750			
; TYPE: DNA			
; ORGANISM: Sorangium cellulosum			
US-09-568-480-1			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
Length:			
Matches:			
Conservative:			
Mismatch:			
Indels:			
Gaps:			

```
Db 51827 GCCGCTCCGAGGCTGATTGGCGCGGTGCTC-----GAGGAGGC 51868
Qy 186 uGlnValThrAlaAspPheAlaIleLeuLeuAspCys----- 200
Db 51869 CGGAGGGTGTCCCGACCTCCAGTCTTATTTGAGTGGTGAAGTTTCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCGGAGCGTATTGACCGGTAAAGAGCTCGCGCTCGAGATCTCTTCCCTGGTGCTC 51988
Qy 205 pPheSerIleCysGluAspValProAlaGlyAspPro----- 217
Db 51989 GTTCGATATGGCGGAGCAATCTATCGAGATTCGCCCATCGCCGTTACTCGAAGCGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaIle 230
Db 52049 CGTCGGCGGTCTCTCGAGTCGGCGCGGTGTAGTACCGTCCGGTATGTTACGAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGAGATCGAGCAGGCGGCGGACACCGCGCGCTCTCCCGGTGTGCTGCC 52168
Qy 250 rProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGAGCGAGPACCATTTCCACGATTTTCTCCGCTCTTCTGTCGCGGAGCA 52228
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAl 284
Db 52229 AAGATTTCGAGATTATCATTTCTGAAGTATGGATCTCTGATGTGCACGAGGACGCG 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCAGGATACGACACATCAGAGGTTTGAGCTCATCGTCGCGCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGl 324
Db 52346 TGGCAGCGCGCAT-----ATAGAGCCACG----- 52370
Qy 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGNAGGCTCTCTG----- 52385
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTTGCTCGCGCCGAGGCTTCTGCTGCTGCTGAGGCGACAGGCGCATCC 52438
Qy 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuLeuAspGluGl 380
Db 52439 GATCTGGTTCGATATCACACG---GGATTGATTGAGGGGTGGCAGAAGTACGAAGATGA 52495
Qy 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCTCTGCTGCTCGGACCTGGTGTGACGTCTTCGCCCGGCT 52555
Qy 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGGCTTTCCGAGACCGCTGATCTGCCAGGCGAGCATCTCCGCGGGGATCTCCGAGCA 52615
Qy 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427
Db 52616 GCAGGTATCTCTCG-----CGCGCGCGGCGCATAGCAGGCGCGCTTGTGA 52663
Qy 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGl 447
Db 52664 CAGTCCGGTAGTGC----- 52679
Qy 447 yAlaAlaLeuAlaGlnArgAlaSerCysTyrGlnIlyStrpGln----- 462
Db 52680 -GCCACCAATCGCGCGCGCGCTGAGTACGAGCAGGAATGGCCGATGCTCGCTGA 52738
Qy 463 ----ValHisArgPheAla----- 467
Db 52739 CGTCGTCCATCGGATGGCGTTGGAGAGGATGTACTTCCACCGCGCGCGCGGACGGT 52798
```

```
Qy 468 -----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGTCCAGGTCGATTGGTACCGGTGAGGCGCTTACGAAGGCGCTCGCTGGAGA 52858
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuLeuAsnAspGluLeuLe 499
Db 52859 TCTGCTCTCTTTCGAAGACACCGGCGAGGTCTGTGGCAGAGGTTCCAGGGGTCCCGCTGCC 52918
Qy 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGCTCAGGCTCTCTGTTTCGCGCGCGGAGCCCGCGGAGAGTGTGTACGCTTT 52978
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCAGCGCAAGACCTATACAGAGGCTCCGCGAGCGCGTCTTCTCTCTCGC 53038
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLeuAlaLe 551
Db 53039 GGGGCTTGGCTCGTGTGATGACCAAGGCGGAGCGGCTCGCTCGCTATCGTGTCT 53098
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560
Db 53099 GGAAGGG-----CGAGGCGAGGCGTGC 53120

RESULT 8
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Alignment Scores:
Pred. No.: 0.704 Length: 68750
Score: 123.50 Matches: 130
Percent Similarity: 31.97% Conservative: 65
Best Local Similarity: 21.31% Mismatches: 234
Query Match: 3.50% Indels: 182
DB: 4 Gaps: 26

US-10-691-383-2 (1-676) x US-09-568-486-1 (1-68750)
Qy 58 SerGluThrTrp-----LysLysValLysLeu 66
Db 51516 AGTGACACCTGGCTGTGGGACGCGCGCTGGAGCGGCGGAGCGCGGAGCGGTG 51575
Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle 86
Db 51576 CCGCTCGAC-----CTGGTGTCTGGCAGCTTCCATGCGAAGTGGGAGGTC 51620
Qy 87 AlathrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla 106
Db 51621 ATGGAGCGCTTCGCGCAGGCGGTACATCATCGGCACTCTCCGCATATGTGAACGTCTTCTGC 51680
Qy 107 SerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln-ThrCy 126
```

Db 51681 GCT-----GCTGGAGAGCGTACACGATAGACGAGTGTG 51713
Qy 126 sThrAsnSerAspAlaLeuAaspProThrAlaProAsnArgAspAsnValAlaPh 146
Db 51714 CTCGTGAGGCTCAATCTCTGTCTACAGGAAGGTTCATCAAGCAGTGAACAC 51773
Qy 146 eAlaSerArgAspAlaAlaArgAGluArgAspGlyThrGlyThrValCysGlnI 166
Db 51774 CTTGTCCGGA-----TCGGCATCCTT- GTAGGGAGCGAGAGCATTTTGTGACTCTCA 51826
Qy 166 eThrAsnGlyLeuThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe 186
Db 51827 GCCGCTCCGAGAGCTGATTGGCGGCTGCTC-----GAGGAGGC 51868
Qy 186 uGlyGlnValThrAlaAspAspPheAlaLeuLeuGluAaspCys----- 200
Db 51869 CGGAGGGGTGTCGCCAGCTCCAGTCTTATTGAGTGTGGTCAAGTTTCCGGGGAACG 51928
Qy 201 -----IleLeu-----AsnGlyAs 205
Db 51929 GCTCCGCGAGTATTGACCGGTAAAGCTCGCGCTCGAGATCTCTCTCCCTGGTGGCTC 51988
Qy 205 pPheSerIleCysGluAaspValProAlaGlyAaspPro----- 217
Db 51989 GTTCGATATGGCGGAGCAATCTATCGAGATTGCGCCCATCGCCGTTACTCGAACGGCAT 52048
Qy 218 -----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
Db 52049 CGTCCGCGGTGTCGTGAGTCGCGCGCGGTGTAGCACCCTCGCGGAATGTTTCAGCAT 52108
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProProValProThrLeuSerSe 250
Db 52109 CTTGGAGATCGAGACGGAGCGCGCGGACACCGCCGCTCTCCCGGTGTCTGCC 52168
Qy 250 rProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla----- 266
Db 52169 TGACCGGAGCGAGTACCATTTACCGATGTTCTCCGCTCTCTCTCCGCGGAGCA 52228
Qy 267 -----ArgAspValProPheMetGlnTyrGlyThrAspGluLeuThrThrAlaAl 284
Db 52229 AAGATTTTCGAGATTATCCATCTCTGAAGTATGGCATTTCTGGATGTCAGCAGGAGCCAGC 52288
Qy 284 aAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAs 304
Db 52289 TGGCCAGGGATACCCACATCAGAGTTTGAGCTCATCTCGCGCCCAATGTC---ATCCA 52345
Qy 304 pGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrG 324
Db 52346 TGGACCCCGCAT-----ATAAGAGCCACG----- 52370
Qy 324 yProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluPr 344
Db 52371 -----GCGAAGCGTCTCCTG----- 52385
Qy 344 oLysGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs 360
Db 52386 -----TCGTTGTCGCGCCGAGCGCTTCTGCTGCTGTCGAGGGCACAGGGCATCC 52438
Qy 360 pGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuLeuAspGluG 380
Db 52439 GATCTGTTTCGATATACCCAG---GGATTGATTGAGGGGTGGCAGAAGTACGAGATGA 52495
Qy 380 uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa 393
Db 52496 TCTTCGTATCGACCATCGCTCTCTGCTCGGACCTGTGTGACGCTCTCGCGCCGGT 52555
Qy 393 lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe 407
Db 52556 AGGCTTTTCGAGACCGCTGAGTCTGCCAGGCGAGCATCTCCGCGGGGATCTCCGAGCA 52615
Qy 407 rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs 427

Db 52616 GCAGTGATCTCTCG-----CGCGCGCGCGGCATAGCAGGAGCGCTTGTGA 52663
Qy 427 pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleG 447
Db 52664 CAGCTCCGGTGAAGTCG----- 52679
Qy 447 yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln----- 462
Db 52680 -GGACCGAATCCCGCGCGCGGTGCAGTACGCGAGGAATGGCCGATGGCTCCGCTGA 52738
Qy 463 -----ValHisArgPheAla----- 467
Db 52739 CGTCTCCATCGGATGCGTTGAGAGAGATGATCTCCACCGCGCGCGCGCAGGT 52798
Qy 468 -----ArgProGluAlaLeuLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 483
Db 52799 TTGGGTCCACGGTCGATTGCGTACCCTGAGGCGCGCTTCCAGAAAGCGCTCGCTGGAGA 52858
Qy 483 pLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLe 499
Db 52859 TCTGCTCTCTTTCGAAAGACACCGCGGAGGTGTCGACAGAGTTCAGGGGCTCCGCTGCC 52918
Qy 499 uLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe 519
Db 52919 GCAGCTCAGGCTTCTGCTTTCGCGCGCGGACCGCGGAAAGAGTGTGTACGCTTT 52978
Qy 519 uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh 539
Db 52979 GGAATGCGACGCAAGACCCATATACAGAGGTCCGCGACGCGCTTCTTCTCTCCGC 53038
Qy 539 r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 551
Db 53039 GCGGGCTTGGCTCGTCTGATGACACAGCGCGGAGAGCGCTGCGCTCGTATCGCTGCT 53098
Qy 551 uIleGlyLeuAspArgGlyGlyGluCys 560
Db 53099 GGAAGGG-----CGAGGCGAGGCGTGC 53120

RESULT 9
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkie, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Alignment Scores: 0.704 Length: 68750
Pred. No.: 123.50 Matches: 130
Score: 31.97% Conservative: 65
Percent Similarity: 21.31% Mismatches: 234
Best Local Similarity: 3.50% Indels: 182
Query Match: 4 Gaps: 26
DB:

US-10-691-383-2 (1-676) x US-09-568-472-1 (1-68750)

; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Alignment Scores:

Pred. No.:	0.704	Length:	68750
Score:	123.50	Matches:	130
Percent Similarity:	31.97%	Conservative:	65
Best Local Similarity:	21.31%	Mismatches:	234
Query Match:	3.50%	Indels:	182
DB:	4	Gaps:	26

US-10-691-383-2 (1-676) x US-09-567-899-1 (1-68750)

QY	58	SerGluThrTrp-----LysLysValIysLeu	66
DB	51516	AGTGCACACCTGCTGTGGACGCCCGCTGGACGGTGGACGGCGCAGACGGCGAGCGGT	51575
QY	67	ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIle	86
DB	51576	CCGCTCGAC-----CTGGTCTCGGACGCTTCATCGCAAGTGGGAGGTC	51620
QY	87	AlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla	106
DB	51621	ATGGAGCGCTCGCGCAGCGTACATCATCGGCACCTCTCCGCATATGGAACGTCTCTGCG	51680
QY	107	SerGlyValPheProGluLeuHisIlyshisThrGlyGluGlyArgHisLeuGln-ThrCys	126
DB	51681	GCT-----GCTGGAGAGCGTCAACAGTACAGCAGATTG	51713
QY	126	SThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPhe	146
DB	51714	CTCGTCAGGCTTCAAATCTCTGCTACAGGAAGGTCATCAAGCGATGATGGAACAC	51773
QY	146	eAlaSerArgArgAspAlaAlaAArgArgGluArgAspGlyThrGlyThrValCysGlnIle	166
DB	51774	CTTGTCGCGA-----TCGGCATCCTT-GTAGGGACGAGGAGCATTTTGTGAGCTCTCA	51826
QY	166	eThrAsnGlyIleuThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLe	186
DB	51827	GCGCTCGCGGACCTGATTGGCGCGGTGCTC-----GAGGAGGC	51868
QY	186	uGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys-----	200
DB	51869	CGGGAGGATGTTTCGCGGACCTCCAGTCCTATTGAGTGGTGCAGATTGTCGGGGGAACG	51928
QY	201	-----lleLeu-----AsnGlyAs	205
DB	51929	GCTCGCGGAGATTGACCGGTAAGACGCTCGCGCTCGAGATCTCTCTCCCTGGTGGCTC	51988
QY	205	pPheSerIleCysGluAspValProAlaGlyAspPro-----	217
DB	51989	GTTTCGATATGCGGAGCGAATCTATCGAGATTCCGCCCATCGCCGCTTACTCGAACGGCAT	52048
QY	218	-----AlaGlyArgLeuValAsnProThrAlaAlaPheAlaIle	230
DB	52049	CGTGGCGGTGTCGTGAGTCTCGCGCGCGGCTGTAGCACCGTCGGGAATGTTTCAGCAT	52108
QY	230	eAspIleSerGlyProAlaPheSerAlaThrThrIleProValProThrLeuSerSe	250
DB	52109	CTTGGAGATCGGAGCAGGACGGCGCGGACCAACCGCCCGCTCTCCCGGTGTGCTGCC	52168
QY	250	rProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla-----	266
DB	52169	TGACCGGAGCGAGTACCAATTTCACCGATGTTTTCTCCGCTCTCTCTGCTCGCGGGAGCA	52228
QY	267	-----ArgAspValProPheMetGlnTyrGlyThrAspGluIleThrThrThrAlaAl	284
DB	52229	AGAAITTCAGATATCCATTCTCGAAGTATGCGCATCTCGATGTCGACGAGGCCAGC	52288
QY	284	aAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGlySerAs	304
DB	52289	TGGCAGGATACGCACATCAGAGGTTTGACGCTCATCTGTCGGGCCAATGTC--ATCCA	52345

Qy	304	pglyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGl	324
Db	52346	TCGACCCCGCAT-----ATAGAGCCACG-----	52370
Qy	324	yProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPr	344
Db	52371	-----GCGAAGCGTCTCTG-----	52385
Qy	344	oLySGlnGluThrPheAlaPro---AspLeuAsnTyrMetValAsp-----PheAs	360
Db	52386	-----TCGTTGCTCGCGCGGAGGCTTCTGGTGTGCTGAGGGCACAGGCATCC	52438
Qy	360	pgLutPrLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGl	380
Db	52439	GATCTGGTTTCGATATACCCACG---GGATTGATTGAGGGGTGGCAGAAAGTACGAGATGA	52495
Qy	380	uLeuArgPhe-----IleArgAsnAlaArgAsp-----LeuAlaArgVa	393
Db	52496	TCTTGTATGACCAATTCGCTCTCTGCTCGGACCTGGTGTGACGTCTTGGCGCGG	52555
Qy	393	lSerPheValAspAsnIle-----AsnThrGluAlaTyrArgGlySe	407
Db	52556	AGCATTGCGGAGCGCGTGAAGTCTCCAGCGCAGGATCTCGCGGGGATCTCCGACA	52615
Qy	407	rLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAs	427
Db	52616	GCAGGTGATCTCTCG-----CGCGCGCGGCATAGCAGGAGCGGTGTGTA	52663
Qy	427	pSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGl	447
Db	52664	CAGCTCCGTGAGTCTG-----	52679
Qy	447	yAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGln-----	462
Db	52680	-GGACCGAATCGCGGCGCGGTGTCAGTACGCGAGGAATGGGCCGATGGCTCGCTGA	52738
Qy	463	----ValHisArgPheAla-----	467
Db	52739	CGTCTCCATCGATGGCGTTGGAGAGGATGTACTTCCACCGCGCGCGCGCAGGT	52798
Qy	468	-----ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs	483
Db	52799	TTGGTCTCCACGATTCGATTTCGTTACCGGTGGAGCGCGCTTCACGAAGCGCTCGCTGGAGA	52858
Qy	483	pLeu-----AspAlaAspPheAspIleSerLeuLeuLeuAsnAspGluLeuLe	499
Db	52859	TCGCTCTCTTTCGAAAGACACCGCGCAGGTGTGTGCAGAGGTTACGGGGCTCCGCTGCC	52918
Qy	499	uLyArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLe	519
Db	52919	GCAGCTCGAGGCTTCTGCTTTTCGCGCGCGGACCCGCGGAAGAGTGGTTGTACGCTTT	52978
Qy	519	uProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaTh	539
Db	52979	GGAATGGCAGCGAAAGACCTTATACAGAGGCTCCGCGCAGCGGCTCTCTCTCTCCGC	53038
Qy	539	r-----GlnAsnGlyAlaPheAlaThrValLeuLysAlaLe	551
Db	53039	GGGGGCTTGGCTGTGTGATGGACAGGGCGGGAGCAGCGCTCGCTCGTGTCTGCT	53098
Qy	551	uIleGlyLeuAspArgGlyGlyGluCys	560
Db	53099	GGNAGGG-----CGAGGCGAGCGCTGC	53120

RESULT 11
US-09-221-017B-21/c
; Sequence 21, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 PAGE MILL ROAD
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows
 SOFTWARE: FastSEQ for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/221.017B
 FILING DATE: 23-DEC-1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1182
 FILING DATE: 31-DEC-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP1546
 FILING DATE: 30-JAN-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PP2911
 FILING DATE: 09-APR-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU98/01023
 FILING DATE: 10-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Monroy, Gladys H
 REGISTRATION NUMBER: 32,430
 REFERENCE/DOCKET NUMBER: 27340-20021.00
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: PORPHYROMONAS GINGIVALIS
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: 1...2843
 US-09-221-017B-21

Alignment Scores:
 Pred. No.: 0.00536 Length: 2843
 Score: 122.00 Matches: 141
 Percent Similarity: 35.52% Conservative: 113
 Best Local Similarity: 19.72% Mismatches: 264
 Query Match: 3.46% Indels: 198
 DB: 4 Gaps: 38

US-10-691-383-2 (1-676) x US-09-221-017B-21 (1-2843)

Qy 31 GlyTrpArgGlnLeuGluGlyGluLysSerLeuGlyPheHisProSerGluThrPro 50
 |||||
 Db 2190 GGTGGTGGGACAAATGAGATCGACGATATTGTCATCTTCGATCGCGCGCGTCT 2131
 |||||
 Qy 51 TyrIleLys-----TyrLeuGluGlySer 58
 |||||
 Db 2130 GCGGTGTGACGCGCAACGTTTGACCGGGGGCTGGATCTGCTCCCTCTTTTGAAGGCGGT 2071
 |||||
 Qy 59 GluThrTrpLysValLysLeu-----Pro 67
 |||||
 Db 2070 CTGAGCTTCTCCTCTGCTCGTCTGTTTCAAGGCGCACTCGATGATATATGCTTCGCGTCCC 2011
 |||||

Qy 68 ThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg----- 83
 |||||
 Db 2010 GATACGGCTCTGAAGATTGAGTTCTCTGGATCTCTGAAAGATTGAAGATCCGGAAGAG 1951
 |||||
 Qy 84 ---ValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAsp 102
 |||||
 Db 1950 CCGAGCCGGCTACGACTCAATATCAATACATCTTGCTCAGGAACCTGC----- 1903
 |||||
 Qy 103 GluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHis 122
 |||||
 Db 1902 CGGCTCTCTGCTCGGATGTTGACAGACCCCAATC-CGACACATTTGGATGCTCGA--- 1847
 |||||
 Qy 123 LeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgAsp 142
 |||||
 Db 1846 ---CTTGAATGCCAAAGTACGTCT-----TTTCCAAGCGACATGACGGACTGAA 1799
 |||||
 Qy 143 AsnValAla-----PheAlaSerArgAspAlaAlaArgArgGluArgAsp 158
 |||||
 Db 1798 AGCTCAGCTGGCCATCTTCTTTTCTCGAAAAAATACCGGTCTCGATCTCAAGA--- 1742
 |||||
 Qy 159 GlyThrGlyThrValCysGln-IleThr---AsnGlyGluThrAspLeuAlaThrMetPh 177
 |||||
 Db 1741 -----ACTGCATCCGATGTCAGTCTCTGTCGCGCACACTTTGGGAGTGGCTCGTCT 1691
 |||||
 Qy 177 eHisLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspPheAlaIleLe 197
 |||||
 Db 1690 TCATATCTCTTTGCCGAATAGTCACATCGGTATGCTCCGGA----- 1648
 |||||
 Qy 197 uGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspPr 217
 |||||
 Db 1647 ----TACTGTGTTCTCTCGCCCGAAACTCTGCTTCCTGCCATAGGGTGGAGCGGACTT 1592
 |||||
 Qy 217 oAlaGlyValGluValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPh 237
 |||||
 Db 1591 GGGCGGACGG--ATCCGTCCTCC-----GACCTGAAAGTTTGTCTCC 1550
 |||||
 Qy 237 eSerAlaThrThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLe 257
 |||||
 Db 1549 GGCACCTCGATACGATGATATATTGAATCTCTCTGTGCCCGAGCTGACAGCTGCTTC 1490
 |||||
 Qy 257 uAlaGluLeuTyrrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrrGlyThrAs 277
 |||||
 Db 1489 CGGCGAA-----GCGGTTCTGCTGAAATCAGT---ATGAACATATGCG---GA 1448
 |||||
 Qy 277 pGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGly----- 292
 |||||
 Db 1447 TGAAGTACCTTCGACGCGCATAGGTTGGCAGAAATAGGAAAAAGGAACGATGCA 1388
 |||||
 Qy 292 ----- 292
 |||||
 Db 1387 TGTCAAGTCCGGTTGGATGAGCTTAAAGCTCCGGAGTCCGGACCTACATGCTGGCAGAACG 1328
 |||||
 Qy 293 ----PhePro-----AsnLeuAspAlaValSerIle----- 301
 |||||
 Db 1327 TCTTTTCCCGATAAGCTCCCACTCAATGCGCCACTGGGTGCTATCGATATATATGCG 1268
 |||||
 Qy 302 -GlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyVa 321
 |||||
 Db 1267 AGCGCTAACGCGCGCTGAC-----GATCTCGCGCTGAAAACCTCGTGTGGAGAC 1217
 |||||
 Qy 321 lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIle-- 340
 |||||
 Db 1216 GGATCGCGAGCTTGGACCATACAG-----GGCAGCATCGCGCTCGATCTCTCTG 1166
 |||||
 Qy 341 -----ThrValGluProLysGlnIleThrPheAlaProAspLeuAsnTyrrMetVa 357
 |||||
 Db 1165 GCGACTTGGCGATGTAGATGCTCATGCGACTACCGAAAGACTGGAACCTCAAGCTCATGCG 1106
 |||||
 Qy 357 lAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLe 377
 |||||
 Db 1105 AGACCGCAGCATGG-----CTGAACATGT 1079
 |||||
 Qy 377 u-----AspGluLeuLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPh 395
 |||||

Db	1078	GGCGTTTCGACTGAAGATGAAGCCCGACTGGCTGCGCGCAATCTCTCTCGGGAAGTGT	1019
Qy	395	evalAspAsnIleAsnThrGluAlaTyrArgGlySerLeu	412
Db	1018	GGAGGGACGGGTGTCGCACCTGACTTTCCGAGGCTATACCTATGAGACCTGACATCGA	959
Qy	412	uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl	432
Db	958	CTTACAGGCC	932
Qy	432	agIlyPheValAsnPheGlyThrSerHis---TyrPheArgLeuIleGlyAlaIleGluLe	451
Db	931	AGGCATCCTGAATATGAACACCCCAATGCCATATCCGGCTATCTCTCCCGAGAGGG	872
Qy	451	uAlaGlnArgAlaSerCysTyrGlnLysTrpGln---ValHisArgPheAl	467
Db	871	GCTGCCTTTCTCTCGAGTCTTCCCGTTTCGAATGGAATCTGACGCGCACAACT--	814
Qy	467	aArgProGluAlaLeu	472
Db	813	-CGTCCCGATCGACTTTTGGCCGGCTTCGGAGTCCCATCGCCCGATGCTCTTCTGATTTC	755
Qy	473	GlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLe	492
Db	754	TTCGGAGAAATGGTCGGAGTGTATGTGACGATCTGACGGAACGATACGAATAGATAC	695
Qy	492	uLeuGlu-----AsnAspGluLeu---LeuLysArgValAlaGluIleAsnAl	507
Db	594	CTGGATTGGCAACCTCGACAAAGAAGCTGCATCTGCACATGTGAACTGGCTATGGC	635
Qy	507	aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro-----GlnAlaIleGlnVa	525
Db	634	AAGGGTGGAGCGACATAGCTCCATTACCCCTCAGTTCGCTTCTCTTGAGGAGCACTGCA	575
Qy	525	lGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAl	545
Db	574	TGGCGAGTTCCACATTTGTCCAAAGTCGCTTCG-----ATGCTCGAAGAGGTGGTTGG	524
Qy	545	aThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlyCysPheProAsnProVa	565
Db	523	CGNCATCTTCTCGACTTGGGGCCCATAGTCCGAGGCAAGA-----	481
Qy	565	lPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla-----	580
Db	480	-GTGCCCGCATGAGATCGCGTCTGCGTTTGGCGGTTCGACGGCTCTTGTGCCAAGGAAGT	422
Qy	581	CysLeuThrTyrGlu-----GlyGluIleAsnLysLeuAlaValAsnValAlaIlePh	597
Db	421	GTGCAGCTTCTTCGGTGGCTTTGGCGGAAGCGGATTTCTATCAAGATCAGGCGAGCCTA	362
Qy	597	eGlyArg-----GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLe	615
Db	361	TTCCGAATTGCTTCGAAGCTCGATTCGCACATTCGTACCGACAGGCTCATG-----	310
Qy	615	uLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeu	629
Db	309	-----ACTATCAACAAAGAGCTA	292

RESULT 12

US-09-103-840A-1
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103, 840A
 ; CURRENT FILING DATE: 1998-06-24

Db 2381258 AGCGCCGACCGTTCGTCGTCGCCCATCAACCGTTCGCGTATTGCGTAACCCACGCTGCC 2381317
Qy 268 AspValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeu 287
Db 2381318 ACGGTGGCCACCGCAGCAGCGCGCGCACGAA-----ACGGCCGCGCGGGTATACG 2381371
Qy 288 AlaGlyMetGlyGlyPheProAsnLeuAsp-----AlaValserIleGlySerAspGly 305
Db 2381372 TCCCATTTGGGGGATGCTTACCTACCTACCGAGTTGGCGGCAACCATGTCATGCACGCG 2381431
Qy 306 ThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPro 325
Db 2381432 GCT-----CTGGTGACCAACCACTTCTCGGTGTCAACACCATCCG 2381473
Qy 326 Phe---ValSerGlnLeuValAsnSerPheThrIleAspAlaIleThrValGluPro 344
Db 2381474 ATCGCCCTCAACGAGCGCGACTACCTGCGCATGTGGATCCAGCGCGCGCACCGTGC- 2381527
Qy 345 LysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTyrLeuAsn 364
Db 2381528 -----ATGAGCCACTATCAAGCGCTGCGGCAC 2381554
Qy 365 IleGlnAsnGlyGlyProPro-----AlaGlyProGluGluLeuAspGluLeu 381
Db 2381555 GA-AAGCGTGGCGCGGACCCACGACGCGCGCGCGCGCATAGTAC---CAGTGC 2381610
Qy 382 ArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThr 401
Db 2381611 GGCCAGCTCGCGGCTAGCAGCAGCTTCCCGACCGCGCAATTTGATCTGTCAGCTACT 2381670
Qy 402 GluAlaTyrArgGlySerLeuIleLeu-----LeuGluLeuGlyAlaPheSer 417
Db 2381671 CAAGGATTTCTGAGCTGCTGCGCTATCTGGCTGTTGAGTGTGCGCGCGCGCTCGG 2381730
Qy 418 ArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPhe 437
Db 2381731 CGACCTCATCGCCAGGT-----GTTGCACTGGTTTCATCTC 2381766
Qy 438 -----GlyThrSerHisTyrPheArgLeuIleGlyAla----- 448
Db 2381767 GTTCGTGTCGGTCCAGTCTTCAGTTCCTCGCTACTCTGTTGCTGCGCGGCTGACCGGCT 2381826
Qy 449 -----AlaGluLeuAla 452
Db 2381827 TTTCGACCGTTCGCCCGCTGACAGTCCGCTCTGTTGCTGCGCGGCTGACCGGCT 2381886
Qy 453 GlnArgAla-----SerCysTyrGlnLys 460
Db 2381887 TGCCGGGCTCGTCCGTTATCGGGCGCGCGGACCAATGTCGACGTCGACTCCGA 2381946
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIle 480
Db 2381947 TGGTCCC--AGCCGGCAAGCTGCGCTCGCGCCACCGGAGTCCACCTGTTGGGTACC-- 2382001
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys 500
Db 2382001 ----- 2382001
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
Db 2382002 -----AACCGGCTCGCTCGTTACACCGCCGCGCACCC 2382034
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 2382035 GCTCCGACCATGTCGGCGGACCCGACCGCATCTCCGATCCCGATCCAGTCCGCGCCAA 2382094
Qy 541 AsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyLysCys 560
Db 2382095 GGC-----CTTTACGGGTGCGTGTGCTCCGAC-----GGGGAAGGG 2382130
Qy 561 PheProAsnProVal 565
Db 2382131 TTC---AACCGGATC 2382142

RESULT 13

US-09-252-991A-15601
; Sequence 15601, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15601
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15601
Alignment Scores:
Pred. No.: 0.0127 Length: 2208
Score: 117.00 Matches: 144
Percent Similarity: 32.99% Conservative: 82
Best Local Similarity: 21.02% Mismatches: 255
Query Match: 3.32% Indels: 206
Gaps: 33
DB:

US-10-691-383-2 (1-676) x US-09-252-991A-15601 (1-2208)

Qy 52 IleLysTyrLeuGluGly-----SerGluThrTrpLysLys 63
Db 92 GTGATTACTTGACGGCGAGTGGACCGACCGATCTCAGCGGCGCACCCTACCGGCG 151
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg 83
Db 152 ACCAACCGACCGCCTACGCGCTCGCGCACCAAGATCTCAGCACCATCAAGAGCAAG 211
Qy 84 -----ValArgIleAlaThrAlaLeuAlaValValLeu 94
Db 212 GAGTACAAAGGCGATCGCGGCAACGAGCTCGCATCGACACACCGCGCGCATCAT 271
Qy 95 AlaAlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPhe----- 110
Db 272 GCGGCG-----CTGATGAGGACCGACGGTGCCAGCGGCTGACCTGGGCTACCTA 322
Qy 111 -----ProGluGluHisLysHisThrGlyGlyArgHisLeuGlnThrCys 126
Db 323 ACCCATCTCGGCGCGGAGCGGCAAGCGCGCGGCGGAGGGCTTCGAACCTGCGCACCC- 379
Qy 127 ThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPhe 146
Db 380 -----GACGAACATGCGCGCTGCGTGGCGGCAAGGG-GCTGCT 417
Qy 147 AlaSerArgArgAspAlaAla-----ArgArgGluArgAsp-GlyThrGlyThrValCy 164
Db 418 GCTGAGCACCGAGGAGCAGCTCAGGCGCGCGCGCGGCATCTCGACCGAGCGCTGCTGT 477
Qy 164 sGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe-----Hi 178
Db 478 GCAGGTGCTGAGGCGCGCTGGAGCTGGCGCGGAGCTACCGGAGCTACCGCGGGAGCA 537
Qy 178 sLysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGl 198
Db 538 CCAGGCGCTGGGCGACGACGCGCGCGCGCAAAACC-----CTCCA 579
Qy 198 uAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal-----ProAlaGlyAs 216
Db 580 GGAAGCGCTG-----CGCGACCTCGGCGCACCGGTGCCAAGCA 615

```
Qy 216 pProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAl 236
Db 616 CGAGTCGGCAAGAGCAACGAGGCAAGCGCGATCGCC-----TTGAGCGTCCGGC 669
Qy 236 aPheSerAlaThrThrIleProProValProThrLeuSerSerProGluLeuAlaAlaG1 256
Db 670 AGGATCGCGCGCGGCGAGCGCGGAGCCTGACCTGGCGCGCGGCGAGCAGCTCGACAG 729
Qy 256 nLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyTh 276
Db 730 CGTGGCGCGGAGAACAGAGAGTGCACCGCGGCGAGAGGTGTGATCAACGCGCGCAG 789
Qy 276 rAspGluIleThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLe 296
Db 790 CGAC-----ATCGGCTGTTCCGCCAGGGCGGC----- 817
Qy 296 uAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAl 316
Db 818 -----GAGCTGGCGCCAGAT 831
Qy 316 aThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheTh 336
Db 832 CACC-----CACCAGGCGCGGATGCTGCTGCAGCGCGGAGAGAACGACATCCG 879
Qy 336 rIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMe 356
Db 880 CTTGGAGACC-----GAGCAGAGCTGGAGTGCAGCGCCAGCCAGCAGCAGCTGCT 930
Qy 356 tValAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluG1 376
Db 931 GGTCAACCGGAGGAGCAGCATCACCTGATGTGGCGGCG----- 970
Qy 376 uLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheVa 396
Db 970 ----- 970
Qy 396 lAspAsnIleAsnThrGluAlaTyr-----ArgGlySerLeuIleLeuLeuGluLe 413
Db 971 -----GCCTACTGACCTCAAGGCGGCGCAACATC-----GAGCT 1005
Qy 413 uGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaG1 433
Db 1006 GGGC-----ATGCCCGGC-----AA 1020
Qy 433 yPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaG1 453
Db 1021 CTTCTGGTCAAGCGCGGAGCAAGCAGTCA-----GTGCGCGCGCCAGCCTCGAAGC 1074
Qy 453 nArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuG1 473
Db 1075 CGAGTCGCGCAGTTCGAGGTGGCGAGAGCGCGCGG-GTTCGTCTCAGCAACTGG 1133
Qy 473 Y-Gly-----ThrLeuHisAsn-----ThrI 480
Db 1134 ACGGCAGACGGCCATGCCCCAACGTCGCTATACCATCACCATGGCCCAACGGCGAGGTCA 1193
Qy 480 lAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuL 500
Db 1194 TCGAGGGCGTCAACGAGCGCGGAGCGCTACCCAGCTGCTGCAGAGGATGCGATGAACA 1253
Qy 500 yArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeu--- 518
Db 1254 TCGCAGAGTTCGATATGAACACAGAAATCGCCGCTCGGCAGTTGCCGGGATAGCAG 1313
Qy 519 -----LeuProGlnAlaIleGlnValGlySer-----ProThrHisProSerTyrProSerG 536
Db 1314 CGGCGGTGGCGCGCGGTTCAGTTCGCGAGCTCCTGAGTGGCGCGCGAGCGCGAGCGCG 1373
Qy 536 lYHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspA 556
Db 1374 GACTGTCTTCAGCGAAGGGGAA-----ATCAGCTTGGCGA 1409
Qy 556 rgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleA 576
```

```
Db 1410 AAGGC-----GTATTCCGCGATTCCATCGACTATTCCACGCTCC 1448
Qy 576 snPhe---GluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn- 594
Db 1449 GCCTGGCGGAGGAGGACTATGTCCTGGCAGGCGCAAGGACTACGTGATGGCCCGCAACG 1508
Qy 595 -----ValAlaPheGlyArgGlnMetLeuGlyIle---HisTyrArgPheAspGlyIle- 611
Db 1509 GCCACATCTACTTCGGGAGGAGTTCGGCGGCGTGGCGGACTGGAGCCTGGAGAGCCTGC 1568
Qy 612 -----GlnGlyLeuLeuLeu-----GlyGluT 619
Db 1569 AGCGCCAGGCTGCTTCAATTCAGAAATGATCCCATGCTGCGCAGCACCAGCATGGCGTCA 1628
Qy 619 hrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheG 639
Db 1629 ACCTCTCTTCTGGTGGGCGCTACCGCAAGCAGGAGGAGTTCTCTTAGGCGACAGTACG 1688
Qy 639 luPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleA 659
Db 1689 CTTACCGTGGAGCGCGGGAAGACGTTGAAG-----GACTACAACATCGAACAGC 1739
Qy 659 spGlyAspMet 662
Db 1740 AGGAGACATC 1750

RESULT 14
US-09-252-991A-15639
; Sequence 15639, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15639
; LENGTH: 3132
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15639

Alignment Scores:
Pred. No.: 0.0226 Length: 3132
Score: 117.00 Matches: 144
Percent Similarity: 32.99% Conservative: 82
Best Local Similarity: 21.02% Mismatches: 255
Query Match: 3.32% Indels: 206
DB: 4 Gaps: 33

US-10-691-383-2 (1-676) x US-09-252-991A-15639 (1-3132)
Qy 52 IleLysTyrLeuGluGly-----SerGluThrTrpLysLys 63
Db 1378 GTGCGATTACCTGGACGCGGACTCGCAGCAGCGATCGTACCAGCGCGCATCTACCGGCG 1437
Qy 64 ValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg 83
Db 1438 ACCAACCGACCGCCCTACGCGCTACCGCGACCAAGATCTCTAGCACCACCAAGAGCAAG 1497
Qy 84 -----ValArgIleAlaThrAlaLeuAlaValValLeu 94
Db 1498 GAGTACAGGGCAGTCGGGCCAAGAGCTGCGATCGACGACACCAGCGCGCAGATCAGT 1557
Qy 95 AlaAlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPhe----- 110
```


Db 1558 GCGGCG-----CTGATGAGCGACACCGTGCCAGCGCGGTGCACCTGGGCTACCTA 1608
Qy 111 -----ProGluGluHiLysHisThrGlyGluGlyArgHisLeuGlnThrCys 126
Db 1609 ACCATCTCTCGGCCGGAAGCGCGCGCGCGGCTTCGAACTGCGCACCC--- 1665
Qy 127 ThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPhe 146
Db 1666 -----GACGAACATGGCGGTGGTGGCGCCAAAGGG-GCTGCT 1703
Qy 147 AlasArghAspAlaAala-----ArgArgGluArgasp-GlyThrGlyThrValCy 164
Db 1704 GCTGAGCACCAGAGAGAGCTCAGGCGCGCGCGCGCCATCTCAGCAGAGCGTGTGGT 1763
Qy 164 sGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPhe-----His 178
Db 1764 CGAGGTCTGGAGCGCGCGCTGGAGCTGGCCCGGAGCTACGCGGGGAGCA 1823
Qy 178 LysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaLeuGlu 198
Db 1824 CCAGGGCGTGGGCGACGCGCGCGCGCGCAACCC-----CTCCA 1865
Qy 198 uAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal-----ProAlaGlyAs 216
Db 1866 GGAAGCCGTG-----CGCGACCTCGGCCACCGTGCCAAACGA 1901
Qy 216 pProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAl 236
Db 1902 CGAGTCCGGCAAGAGCAACGAGGCAAGCGCGCATCGCC-----TTGAGCGTTCGGC 1955
Qy 236 aPheSerAlaThrThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlu 256
Db 1956 AGGATCGCGCGCGCGAGCGCGCGAGCTGACCTGCGCGCGCGCGAGCAGCGTCGACAG 2015
Qy 256 nLeuAlaGluLeuTyrrPheAlaLeuAlaArgAspValProPheMetGlnTyrrGlyTh 276
Db 2016 CGTGGCGCGCGAGAACACAGAGGTGACCGCGCGCGAGAGAGGTGTGATCAACGCGCGCAG 2075
Qy 276 rAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLe 296
Db 2076 CGAC-----ATCGCGCTGTTCCGCCAGGGCGGC----- 2103
Qy 296 uAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAl 316
Db 2104 -----GAGCTGCGCCAGAT 2117
Qy 316 aThrPheValGlyValGluThrGlyProPheValSerGlnLeuValAsnSerPheTh 336
Db 2118 CACC-----CACCAGGCGCGCGATGCTGCTGCGCGCGCAGAGAACGACATCCG 2165
Qy 336 rIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrrMe 356
Db 2166 CTTGGAGACC-----GAGCAGAGCGTGGAGTCTGCGCGCGCAGCAGCAGCGTGTCT 2216
Qy 356 tValAspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGlu 376
Db 2217 GGTCAACCGCAAGGAGCACATCACCTGATGTGGCGCGC----- 2256
Qy 376 uLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheVa 396
Db 2256 ----- 2256
Qy 396 lAspAsnIleAsnThrGluAlaTyrr-----ArgGlySerLeuIleLeuLeuGluLe 413
Db 2257 -----GCCTACTGACCTCAAGGCGGCAACATC-----GAGCT 2291
Qy 413 uGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerArgGlnAlaGlu 433
Db 2292 GGGC-----ATGCCCGGC-----AA 2306
Qy 433 yPheValAsnPheGlyThrSerHisTyrrPheArgLeuIleGlyAlaAlaGluLeuAlaGlu 453
Db 2307 CTTGTGTCTAAGCGCGGAGACACATGTCAT-----GTGCGCGCGCGCGCTGGAAGC 2360

Qy 453 nArgAlaSerCysTyrrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlu 473
Db 2361 CGAGCTCCGCGAGTTCCAGGTGGCGAGACGCGCGCG-GTTCTCTCTCAAGCAACTGG 2419
Qy 473 Y-Gly-----ThrLeuHisAsn-----ThrI 480
Db 2420 ACGGCGAGCGCGCATGCCCAACGTGCCCTATACCATCATCCATGCCCAACGCGAGGTCA 2479
Qy 480 lAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuL 500
Db 2480 TCGAGGGCGTCCAGCAGCGCGAGCGCTACCCAGCTGCTGCAGAAAGATCCGATGAACA 2539
Qy 500 ysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrrLeu---- 518
Db 2540 TCGCAAGGTTCGATATGAAACACACAGAAATCGCCCGCTCCGCAAGTTCGCGGATAGCAG 2599
Qy 519 -----LeuProGlnAlaIleGlnValGlySer-----ProThrHisProSerTyrrProSerG 536
Db 2600 CGCGCGTTCGCGCGCGGTTCGAGTCGCAAGTCTCTGAGTGGCGCCGACGCGCGAGCGCG 2659
Qy 536 lYHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspA 556
Db 2660 GACGTGCTCTCAGCGAAGGGAA-----ATCAGCTTGGCGA 2695
Qy 556 rgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuLeuA 576
Db 2696 AAGGC-----GTATTGCGCGATTCCATCGACTATTTCACGGTCC 2734
Qy 576 snPhe-----GluGlyAlaCysLeuThrTyrrGluGlyGluIleAsnLysLeuAlaValAsn- 594
Db 2735 GCCTGCGCGACGAGGACTATGTCCTTGGCAGGCAAGGACTACGTGATGCGCGCAACG 2794
Qy 595 -----ValAlaPheGlyArgGlnMetLeuGlyIle-----HisTyrrArgPheAspGlyIle- 611
Db 2795 GCCACATCTACTTCGCGCAGGAGTTCGCGCGCTGGCGGACTGGAGCCTGGAGAGCCTGC 2854
Qy 612 -----GlnGlyLeuLeuLeu-----GlyGlu 619
Db 2855 AGCGCCAGCGTCTGTTTCATTTCAGAAATGACCCATGCTGCGCAGCACCAGCATGGCGTCA 2914
Qy 619 hrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheG 639
Db 2915 ACGTCTCTCTGTTGGGGGCGCTACCCAGCAAGCAGGAGTTTCTCTTCTGAGCGACAGTACG 2974
Qy 639 luPheArgLeuPheThrGlyGluValIleValLeuPheGlnAspGlyThrPheSerIleA 659
Db 2975 CCTACCGGTTGGAGCGCGGAAGAGAGTTGAAG-----GACTACAACATCGAAGCAGC 3025
Qy 659 spGlyAspMet 662
Db 3026 AGGAGAGACATC 3036

RESULT 15
US-09-266-965-76/c
; Sequence 76, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07

! NUMBER OF SEQ ID NOS: 145
! SOFTWARE: Fast-Seq for Windows Version 3.0
! SEQ ID NO 76
! LENGTH: 53500
! TYPE: DNA
! ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Alignment Scores:
Pred. No.: 2.46 Length: 53500
Score: 117.00 Matches: 151
Percent Similarity: 32.48% Conservative: 64
Best Local Similarity: 22.81% Mismatches: 207
Query Match: 3.32% Indels: 240
DB: 4 Gaps: 36

US-10-691-383-2 (1-676) x US-09-266-965-76 (1-53500)

```
Qy 2 LeuCyHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeuArg 21
Db 17507 CTGGGACACGCTCTCGATGTACTTCGAGGACGAGATCACCTCGACGAGCTGTGGCCGC 17448

Qy 22 -----LeuLeuThrSerGluInArgAlaLysGlyTrpArgArgIn----- 35
Db 17447 TCCACACCTACTCCCGCCCGCATCCGCGAGATGCTCAGCGACGCGCAAGCCGACTACCT 17388

Qy 36 -----LeuGlu-GlyGluLysSerLeuGlyPheHisProSerGluThrPr 50
Db 17387 CCAGGGGGCGCATGACTGAACAGCGCACCGGTACGAGGCGGACACCCCTCTCGGGCGC-- 17330

Qy 50 oTyrIleLysThrLeuGluGlySerGluThrTrpLys-LysValLysLeuPro---ThrA 69
Db 17329 -----CGTGGACGGGTGCGGTCCGCGCGGATACCGCTCCCGGGGACCG 17286

Qy 69 spGlyIleSerAlaSerLysIleLeuGlyLysIleMetAla-gValArgIleAlaThrA 89
Db 17285 AT-----GTGAACCTCCCTGCTGGAGCAGGTGCTGACGCGCGCG 17247

Qy 89 laLeu-----AlaValValLeuAlaAlaPro----- 97
Db 17246 CACAGCTCAAGCGAGCGGTGGAGGCGGCGGTGGCCCTGCGCGCGCCCAACAGCC 17187

Qy 98 -----CysLeuAlaPheAspGluValThrAlaSerGlyVal----- 109
Db 17186 CGCGGTGGTGGTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17127

Qy 110 -----PheProGluLysLys----- 115
Db 17126 AACCCGACACCCCGGCGCGGACCGCGCGGTCTGTCGCGGTGCGCGGGCGGCGAGGT 17067

Qy 116 -----HisThrGlyGluGlyArgHisIleLeuGlnThrCysThrAsnSerAspA 131
Db 17066 CCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17010

Qy 131 sp-----AlaLeuAspProThrAlaProAsnArgAspA 143
Db 17009 GAGAACCTCCGGGCGCCCGCCCGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16950

Qy 143 snValAlaPheAlaSerArgAspAla-----AlaArgGluArgAspG 159
Db 16949 GCAAGCTCTGTCGCGCGGACGAGGAGAGACCTGTCTCGCGGAGGCGCGCTTACCGCGACG 16890

Qy 159 lyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeu-----AlaThrMetP 177
Db 16889 GGGTCTGGG-----CTGACGGGAGAGGACACCTGTCTCTCTCTCTCTCTCTCTCTCTCT 16842

Qy 177 heHis-----Lys 180
Db 16841 CCCACGGGTACCGCTGGGTGGTGTTCGCGGAGCTGCTCAGCGGTGCGCGGTGGCGCC 16782

Qy 180 erLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspC 200
Db 16781 CCGTACCGCGGACCGCGCTCTGGCGCGCATCGCGCGGAG----- 16744
```

```
Qy 200 ysIleLeuAanglyAspPheSerIleCysGluAspValProAla----- 214
Db 16743 -----CTGTCCGGTGGTGGCGACCGTGGTGGCCCTGTGTGTCACAGTGGCCCGGCTGTGG 16689

Qy 215 -----GlyAspProAlaGlyArgLeu---ValAsnProThrAlaAlaP 228
Db 16688 CGACCCGGCGGCTGGCGGAGCAGCGCGCGCGCGCGCGCGCTCCCGGTCTCCCGCTCCGCGC 16629

Qy 228 heAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThrL 248
Db 16628 TGGCCATGGTGGTGGCGCGCGGTG----- 16603

Qy 248 euSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgA 268
Db 16602 -----GACGAGCAGCTGGACCGCGCTTACCCAGGAGGTTCGGGACCGGTCTCGCCGCA 16548

Qy 268 spValProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuA 288
Db 16547 AC-----TACGGT-----TCCACGAGACCGGCGCGCTGCTCG 16515

Qy 288 laGlyMetGlyGly-----PheProAsnLeuAspAlaV 299
Db 16514 CCGGACCGCGCGGCTGGAGCCCTTGTGCGCGGTCTCCCTCGCGGGGTGGAGTGGC 16455

Qy 299 alSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheV 319
Db 16454 AACTGACCGCGCGGAGGCGGTGTCGCGCC----- 16423

Qy 319 alGlyValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspA 339
Db 16422 -----GCCGCGACCCCGCGCGC 16407

Qy 339 laIleThrValGluProLysGlnGluThrPheAla---ProAspLeuAsnTyrMetValA 358
Db 16406 TGTGAGCGTACCGGTTCGACGCGCGCGTACGCCATGGCGCATCTCCCGTGGCCCTGTC 16347

Qy 358 spPheAsp-----GluTrpLeuAsnIleGlnAsnGlyGlyProProA 372
Db 16346 CCGGGGCGCTCGCATCTCTGGGACGCGAGACCGCGCGGTCTCCCGGGGCGCGCTGGG 16287

Qy 372 laGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArgAspLeuAlaA 392
Db 16286 TCTCCCGCTGGAGATCGAGAGGTGCTGCGCGTATCCCGGAC----- 16243

Qy 392 rgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGly-----S 407
Db 16242 -----GTGGTGAATGTGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 16194

Qy 407 erLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsn----- 422
Db 16193 ACGGATCGTCCGCGAGGTCTCGCG-----CGCGGCGCGCGCGCTCACCCCGGCGCTGC 16137

Qy 423 -----GlyProPheIleAspSerAsp----- 429
Db 16136 GCGAGACCGCGCGGAGCTGGCGCGCACAGGTGCGCGCGAGTTCGTCTCTCGCGG 16077

Qy 430 -----ArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgL 445
Db 16076 AGAGCTTCGCGTCAACGCGCGCGGCAAGGTGCGG-----GCCGCGCTCGCTTACCGCC 16023

Qy 445 euIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnVal---- 463
Db 16022 TCACCGGAGCGG-----GCCGAGCGCGCGCGCGGTACAGGATCCGAAGTGTCT 15969

Qy 464 -----HisArgPheAlaArgProGluAlaLeuGly-----GlyThrLeuHisAsnThrI 480
Db 15968 TCGCGCTGACGAGCTTGGGCGCGCTGAGGCGACTCGCCAGGCTGCGCGCGCGCTCTCTC 15909

Qy 480 leaIaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuL 500
Db 15908 TCAGCGGGAGGTGGGTGGCGCGCGATCCCTG-----GAGTGGCTGC 15864
```

Qy	500	ysArgValAlaGlu-----IleAsnAlaAlaGlnAsnProAsnAsnGluV	515
Db	15863	TGCGCAGCGCCACCGCTCTGGGGGTGTGACCAACCGGGGGCAAGAGCCCGGGGACCGGG	15804
Qy	515	al	515
Db	15803	TC	15802

Search completed: September 19, 2004, 10:15:25
Job time: 3134.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: September 19, 2004, 03:56:03 ; Search time 869.364 Seconds
(without alignments)
3303.314 Million cell updates/sec

Title: US-10-691-383-2

Perfect score:

Sequence: 1 MLCHAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVAQCQA 676

Scoring table: BLOSUM62

scoring table: `prosumez`
`Xgapop 10.0` , `Xgapext 0.5`

```

"qapop 10.0 ;
"qapext 0.5

```

```

:3app 6.0 , Fqapext 7.0
:3app 6.0 , Fqapext 7.0

```

Delop	6.0	Delest	7.0
-------	-----	--------	-----

Searched: 3373863 reqs. 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB sea length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 2000000000

Post-processing: Minimum Match 0.9

POST-PROCESSING: MINIMUM MATCH 0%
MAXIMUM MATCH 100%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```

Command: cd /home/linne/parameters;
-MODEL=framet+gzn.model -DEV=xlh
-Q/cg2_1/USPTO.spool/QUS10691383/runat.170920024.102655.1733/app_query.fasta_1.1230
-D=Ns Geneseg 25Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human04.cdi
-LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691383 -CGCN_1.653 -@runat.170920024.102655.1733 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCGPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELOPEXT=7

```

Database : N Geneseq 29Jan04 : *

```
Database : N_Geneseq_230allv4:
1: geneseqm1980s:*
```

1: geneseqm1990s:*

2: **geneeqn200s:**
3: **geneeqn200s:***

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseg2002s: *

7: geneSeqn2003as:*

8: geneseqn2003bs:*

9: geneSeqn2003cs:*

10: geneseqn2004s:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB	ID	SUMMARIES	
						Query	Description
1	3528	100.0	2931	3	AA10305		Fucus van
2	3528	100.0	2931	6	AA28155		Fucus dis
3	626	17.7	1797	4	AA11050		Corallina
4	623	17.7	1794	2	AAV56020		Haloperox
5	614.5	17.4	1791	2	AAV56021		Haloperox
6	131	3.7	9579	7	AAV54222		Streptomy
7	131	3.7	52101	7	AAV54217		Streptomy
8	129.5	3.7	11679	8	ADA30090		DNA encod

9	124	3.5	18930	7	ACA43739	Prokaryote
10	123.5	3.5	68750	3	Az55887	Sorangium
11	123	3.5	1962	7	ACF70799	Phototroph
12	123	3.5	5820	6	AB578676	A. orient
13	123	3.5	110000	7	ACF67367	Continuation (39 o
14	123	3.5	110000	7	ACF65388	Continuation (10 o
15	118	3.3	3426	4	AA61025	AA61025 P. putida
16	118	3.3	349980	6	ABQ81847	Bifidobac
17	117.5	3.3	3505	7	ACA35977	Prokaryot
18	117.5	3.3	110000	4	AA199682	Continuation (24
19	117	3.3	3456	2	AA785397	Pseudomon
20	117	3.3	53500	3	AA55842	Complete
21	117	3.3	53500	9	ADE10261	S. lavend
22	116.5	3.3	3971	7	AA4747845	Haemophil
23	116.5	3.3	11883	3	AA10263	Bordetell
24	116.5	3.3	12036	6	AA04668	FHA struc
25	116.5	3.3	19547	4	AA59601	Propionib
26	116.5	3.3	19547	7	ACF64530	Propionib
27	116.5	3.3	30121	2	AA742063	Continuation (19 o
28	116.5	3.3	110000	2	AA742063	Continuation (18 o
29	115.5	3.3	2271	2	AA06844	Amylase g
30	115	3.3	65140	4	AA017184	Streptomy
31	115	3.3	125401	4	AA017186	Streptomy
32	114	3.2	47988	9	ADE86070	Streptomy
33	114	3.2	349980	6	ABQ81845	Bifidobac
34	113.5	3.2	2824	6	ABN89350	Chicken N
35	113.5	3.2	2824	6	ABN89351	Chicken N
36	113.5	3.2	10419	3	AA55188	Cenarchae
37	113.5	3.2	42432	3	AA455187	Cenarchae
38	113	3.2	88421	6	AA140781	88421nt g
39	112	3.2	2520	2	AAQ40580	Glutaryl-1
40	111.5	3.2	75236	7	ABV75557	Saccharo
41	111	3.1	25801	4	AB111177	Pseudomon
42	110.5	3.1	4172	4	ABE13509	Drosophil
43	110.5	3.1	110000	6	ABQ67196	Continuation (4 of
44	110.5	3.1	110000	6	ABQ67196	Continuation (5 of
45	110.5	3.1	110000	6	ABQ69245	Continuation (29 o

ALIGNMENTS

RESULT 1	
AAAL0305	
ID AAAA10305 standard; cDNA; 2931 BP.	
XX	
XX AAAA10305;	
AC XX	
DT DT 03-JUL-2000 (first entry)	
XX XX	
DE DE Fucus vanadium bromoperoxidase cDNA.	
XX XX	
KW KW Vanadium bromoperoxidase; vanadium halo-	
KW o-dianisidine oxidation; vanadium ion; l	
KW phenolic adhesive production; ss.	
XX XX	
OS OS Fucus sp.	
XX XX	
FH FH Key Location/Qualifiers	
FT CDS 228..2258	
FT FT /*tag= a	
FT FT /product= "Fucus vanadi-	
XX XX	
PN PN WO200015771-A1.	
XX XX	
PD PD 23-MAR-2000.	
XX XX	
Pf Pf 27-AUG-1999; 99WO-US019893.	
XX XX	
PR PR 10-SEP-1998; 98US-00151189.	
XX XX	
PA PA (REGC) UNIV CALIFORNIA.	
XX XX	
PI PI Vreeland V, Ng XL;	

Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProhenAsnGluValThrTyrLeuLeuPro 520
 Db 1728 CGTGGCGGAGATAAATCGCGCGAGAAATCCCAACAGAGGTACCTACTCTTCCA 1787
 Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
 Db 1788 CAAGCTATCCAAAGTGGGATGCGCAACGACCCCTTCTACCGTCCGCGCACGCTACCCAA 1847
 Qy 541 AsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
 Db 1848 AATGGAGCATTTGCCACAGTTCTGAAGCCCTCATTTGGCTAGATCGGGAGGTGAGTGC 1907
 Qy 561 PheProAsnProValPheProSerAspGlyLeuGluIleLeuAsnPheGluGlyAla 580
 Db 1908 TTCCCTTAACCCCGGTTCCTCAAGCGATCAGCGCTGGAACTTAATCAACTTCGAAGGGGCA 1967
 Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
 Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCCGGTCAACGTTCGATTTGGGAGGCAG 2027
 Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
 Db 2028 ATGCTGGGCATCCACTATCGGTTGCGGTATCCAGGCCCTACTTCTCGAGAGACATC 2087
 Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
 Db 2088 ACTGTACGAACACTTCACAGGAGCTGATGACGTTCCCGAGGAAGCCACCTTTGAATTC 2147
 Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
 Db 2148 CGCTATTATCCAGGAGAGGTGATCAAACTTTTCAGGACGGGACATTTCCATCGATGGA 2207
 Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
 Db 2208 GATATGTGTTCCGTTTGGTTTACACTGGCGTGGCGACTGCCAGGCT 2255

RESULT 2
 AAD28155
 ID AAD28155 standard; cDNA; 2931 BP.
 XX
 AC AAD28155;
 XX
 DT 07-AUG-2003 (revised)
 DT 01-MAY-2002 (first entry)
 XX
 DE Fucus distichus vanadium haloperoxidase encoding cDNA.
 XX
 KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
 KW industrial catalysis; enzyme; catalyst; ss.
 XX
 OS Fucus distichus.
 XX
 FH Key Location/Qualifiers
 CDS 228..2258
 FT /*tag= a
 FT /product= "vanadium haloperoxidase"
 FT
 FN WO200200838-A2.
 XX
 XX 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018602.
 XX
 XX 19-JUN-2000; 2000US-00596794.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Vreeland V;
 XX
 XX WPI; 2002-154651/20.
 DR P-PSDB; AAE17501.
 XX
 XX

PT Isolated nucleic acid for, e.g. to halogenate various substrates
 PT including proteins, has polynucleotide sequence encoding vanadium
 PT haloperoxidase polypeptide with catalytic helical frame.
 XX
 XX Claim 4; Page 49-53; 56pp; English.
 XX
 CC The invention relates to nucleic acid encoding vanadium haloperoxidase
 CC particularly vanadium bromoperoxidase polypeptide which consists of a
 CC catalytic helical frame that complexes a vanadium ion and catalyses the
 CC oxidation of o-dianisidine (ODA). The invention also provides
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the
 CC invention can be fused to other proteins to allow quantification or
 CC localisation of the linked protein. They can be used to halogenate
 CC various substrates including proteins; to produce epoxides from alkenes,
 CC halogenated ketones from alkyne; to produce alpha, gamma-halohydrins
 CC from cyclopropanes and to produce dihalogenated products from alkenes and
 CC alkynes. They can also be used in signal generating systems in place of
 CC horseradish peroxidase; as a component in assays; as enzymatic
 CC antimicrobial agents; and in the production of phenolic adhesives. The
 CC polypeptides can be used in industrial catalysis in a variety of
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
 CC reactions. The present sequence is Fucus distichus vanadium
 CC haloperoxidase encoding cDNA. (Updated on 07-AUG-2003 to correct OS
 CC field.)
 XX
 SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2931
 Score: 3528.00 Matches: 676
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-691-383-2 (1-676) x AAD28155 (1-2931)

Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
 Db 228 ATGCTTTGCCATGCGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTGCTT 287
 Qy 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyLys 40
 Db 288 CGGTTGCTCAGATCAGACGCGCTAAAGTTGGAGAGCCAGCTTAGAGGGGGAGAAA 347
 Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
 Db 348 TCACTAGGTTTTCATCCAGCGAGACGCTTATATCAAGTACTTTGGAAGGCTCTGAGACT 407
 Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
 Db 408 TGAAGAAGGTAAAGCTTCCACGCGCATATCGGCTTCCAGATCTCGGTAAAAATT 467
 Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
 Db 468 ATGGCCAGGGTCCGCATCGCTACCGCTTGGCTTGGTACTTGGCCGCCACCTGTTGGCA 527
 Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGlyGly 120
 Db 528 TTCAGCAGGTCACAGCCAGTGGTGTGTTTCCCTGAGGAACAACAGCACCCGGGGAGGGA 587
 Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
 Db 588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCGCTGGATCCGACGGGCCGAATAGA 647
 Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
 Db 648 AGGACAAACGTAGCTTTTGGCTCGCGCGCGGATGCCGCCAGGCGAGAACGTCGCGGACA 707
 Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
 Db 708 GGGACTGTCTGCCAAATCACTAAACGGGAGAACTGATTTGGCTTACCATGTTCCCAAGTCT 767

```
QY 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
DB 768 CTGCCACACGATGAACTGGGACAGTAACCGCAGACGACTTCGCTATCCTCGAGGACTGC 827
QY 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
DB 828 ATCTTAAACCGAGATTTCAGCATTTGCGAGACGTCCTCGCGAGACCCCGGGGTGCG 887
QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
DB 888 CTGGTCAATCTCCTACCGCTGGCTTGGCCATGACATATCCGGTCCCGCATTTCTCGGTACG 947
QY 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
DB 948 ACAATACCCCGGTACCTACCTTCTCTCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007
QY 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
DB 1008 TACTGATGCGCTGGCGCAGGATGTATCCCTTTATGCAGTATGGCACCGACGAAATTTACC 1067
QY 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
DB 1068 ACTACCCCGGAGCAACCTCGCTGGAAATGGAGGCTTCCAAATCTGGACCGCGTGTG 1127
QY 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
DB 1128 ATAGGTTCCGATGGTAGCGTGGACCCGTTCTCCAGAGCTTCCGAGGACCTTCGTGGT 1187
QY 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
DB 1188 GTTGAAACCGGGCCCTTGTCTCTCAGCTGCTCGTGAACAGCTTCCACCATCGACGCTATT 1247
QY 341 ThrValGluProIysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
DB 1248 ACGTTCGAACCGAAGCAGGAGACATTCGCCCGGACTTGAACATATATGGTCGATTTGAC 1307
QY 361 GluTrpLeuAsnIleGlnAsnGlyGlyProAlaGlyProGluGluLeuAspGluGlu 380
DB 1308 GAATGGCTGAACATTGAGATGGTGGACCCCGCGCGGCCCGCCGAGATTAGACGAGAG 1367
QY 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
DB 1368 CTGCGTTTATCCGTAAACCGCCGACCTGCGCAGGCTCTCTCGTGGCAATATCAAC 1427
QY 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
DB 1428 ACCGAAGCTTATCGCGGTCTCTTATCTCTAGCTTGGAGCTTTCAGCAGGCGCGGT 1487
QY 421 IleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
DB 1488 ATCAACGGTCCATTATCGACAGTATCGGACGGGGCTTCGTCAACTTCGGCACGTCT 1547
QY 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIys 460
DB 1548 CACTACTTCAGATTGATAGTGGCGCGAGCTGCGCGAGGCTGCTCGTTACCAAAAG 1607
QY 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrIleAsnThrIle 480
DB 1608 TGGCAGGTGATCATGTTGACGCGCCGAGGCTCTCGGGGGTACCTCCCAACACCATC 1667
QY 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuIys 500
DB 1668 GCGGGGATCTAGATGACAGACTTCGACATCTCCCTCTTGGAAATGATGAGCTCTTGAAA 1727
QY 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuPro 520
DB 1728 CGTGTGGCGAGATAAATGCGCGCAGAAATCCCAACACGAGGTACCTACTCTTCTCCA 1787
QY 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
DB 1788 CAAGCTATCAAGTGGATGCGCAACGACCTTCTTACCGCTCCGCGCACGCTTACCACAA 1847
QY 541 AsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
```

```
DB 1848 AATGGAGCATTTGCCAGATTCTGAAGGCCCTCATTGGCCTAGATCGGGAGGTGAGTGC 1907
QY 561 PheProAsnProValPheProSerAspAspGlyIleuGluLeuIleAsnPheGluGlyAla 580
DB 1908 TTCCTTAACCCCGTGTTCCTCAAGCATGACGGCTCGAATCAATCAACTTCGAAGGGGCA 1967
QY 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAACGTCCGATTTGGAGGCGAG 2027
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620
DB 2028 ATGCTGGGATCCACTATGCTTTCGACGATATCAAGGCTATCTTCGAGAGACAATC 2087
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 2088 ACTGTACGAACACTTCCACGAGGAGCTGATGACGTTCCGCCGAGGAAGCCACTTTGAATC 2147
QY 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
DB 2148 CGTTTATTCACGAGAGGTCAATCAAACTTTTCAGGACGGGACATCTCTCATCGATGGA 2207
QY 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
DB 2208 GATATGTGTTCCGTTTGGTTTACACTGGCGTGGGAGCTGCCAGGCT 2255
RESULT 3
AAD11050
ID AAD11050 standard; cDNA; 1797 BP.
XX
AC AAD11050;
XX
DT 24-SEP-2001 (first entry)
XX
DE Corallina officinalis vanadium bromoperoxidase (cVBPO) cDNA.
XX
KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
KW signal generation; analytical system; anti-microbial agent;
KW industrial application; biological material; avidin; metabolite;
KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
KW vitamin; intact cell; drug; narcotic; ss.
XX
OS Corallina officinalis.
XX
FH Key Location/Qualifiers
CDS 1..1797
FT /*tag= a
FT /product= "Corallina officinalis vanadium bromoperoxidase
FT (VBPO)"
FT primer_bind complement(1..21)
FT /*tag= b
FT /bound_moiety= "PCR forward primer (AAD11054)"
FT primer_bind 1150..1170
FT /*tag= c
FT /bound_moiety= "PCR reverse primer (AAD11055)"
FT misc_binding complement(1420..1449)
FT /*tag= d
FT /bound_moiety= "PCR probe (AAD11053)"
FT primer_bind complement(1420..1442)
FT /*tag= e
FT /bound_moiety= "PCR upstream primer (AAD11051)"
FT primer_bind 1642..1664
FT /*tag= f
FT /bound_moiety= "PCR downstream primer (AAD11052)"
XX
XX WO200153494-A2.
XX
PD 26-JUL-2001.
XX
PF 24-JAN-2001; 2001WO-US0002383.
XX
XX 24-JAN-2000; 2000US-0177764P.
XX
PR
```



```

Qy 611 IleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630
Db 1675 TTCAGTCAATTCGCTTGGTACAGGTTGCGATTGGAATCTTGGAGAGCAGAGTCTG 1734
Qy 631 ThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1735 ACGTATCGCGAGAACTCTTCTTCAACTTCCGGAAGTTTGATGGAACACTACATCCAGATC 1794

RESULT 4
ID AAV56020
AC AAV56020;
XX 09-DEC-1998 (first entry)
DE Haloperoxidase, CP.BP01 encoding cDNA.
XX Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
XX bromine; iodine; ds.
XX Corallina pilulifera.
XX Key Location/Qualifiers
XX CDS 1..1794
XX /*tag= a
XX /product= "Haloperoxidase, CP.BP01"
XX /note= "the stop codon is not indicated"
XX JF10248581-A.
XX
XX 22-SEP-1998.
XX 06-MAR-1997; 97JP-00070539.
XX 06-MAR-1997; 97JP-00070539.
XX (SAKA ) OTSUKA PHARM CO LTD.
XX WPI; 1998-560733/48.
XX P-PSDE; AAW80550.
XX
XX New haloperoxidase gene - and corresponding vector, transformed host cell
XX and method of preparation.
XX
XX Claim 2; Page 13-15; 20pp; Japanese.
XX
XX This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
XX encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
XX pilulifera. A host cell transformed with a vector containing the
XX haloperoxidase genes can be used for the recombinant production of the
XX enzyme. The haloperoxidases of the invention are vanadium-dependent and
XX have halide specificity of bromine and iodine
XX
XX Sequence 1794 BP; 395 A; 481 C; 505 G; 413 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.19e-48 Length: 1794
Score: 623.00 Matches: 191
Percent Similarity: 44.41% Conservative: 75
Best Local Similarity: 31.89% Mismatches: 233
Query Match: 17.66% Indels: 100
DB: 2 Gaps: 20

US-10-691-383-2 (1-676) x AAV56020 (1-1794)
Qy 141 ArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGCGCCAGAGGCTTCATTCGATACCGCTGTAGCTGCGCGCGAGCTTGCACCTCAAC---CGC 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
Db 88 GGAGTTGTACCATCGTTTGGCAATGGGGAAGAGCTTCTCTACCGCAACCGGACCCCTGAC 147

```

```

Qy 171 ---ThrAsp-----LeuAlaThrMetPheHisSerLeuProHisAspGluLeuGly 187
Db 148 AACACTGATCCGAGCTTCATCGTAGCTTCACAAAGGGTCTTCGATCAGACGAATGGC 207
Qy 188 Gln---ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 208 GCTATTATCGACCCCGACGACTTCTTGGCCCTTCGTCGTCATCAATAGTAGCGGATGAA 267
Qy 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 AAGGAGATCGCCGACCTCACAATTGGGGCCAGCTCGCGACCCGAGACTGCTTACCAATC 327
Qy 222 -----ValAsnProThr 225
Db 328 TGGCGCTCGGATCTGGCGAATTCTCTCGAGCTCGAAGTGGGAGATGGGAGACAGCTCT 387
Qy 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245
Db 388 GCCGCTCTCACCCTCGACCTGGAGGCGCGACGCGCAGTCGATTGCCATGCCACCGCG 447
Qy 246 ProThrLeuSerSerProGluLeuAlaIleGlnLeuAlaGluLeuThrTrpMetAlaLeu 265
Db 448 CTTGTGCTCAGAGCCCTGAGCTCGTGGCGGAGATAGCAGAGCTGTACTGATGGCGCTT 507
Qy 266 AlaArgAspValProPhe-----MetGlnTyr 274
Db 508 GGACGGCAATCGAGTTTCAGCGAGTTGATCCCAAGAACGAGAGATATATTCAGTTT 567
Qy 275 GlyThrAspGluIleThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 568 GCTATTGATCAGCTTAACGGGCTGGAGTGGTTCAACACACCGGCAAGCTCGGAGATCCG 627
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
Db 628 CTTGCGGAAATCGCTCGCCCTCGCGGTGAGGTGACTGTT-----GGAAACTTGTTC 678
Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeu----- 331
Db 679 CGCGGTATTCTCCAGGCTCTGAGGTGCGGCCGCTACCTCAGCCAGTACATCATCGTTGGT 738
Qy 332 -----ValAsnSerPheThrIle----- 337
Db 739 AGCAAGCAGATTGCTCAGCGACAGTTGGTGAACAAACTCTCGTAGAGCCCAATGCTGCT 798
Qy 338 -----AspAlaIleThrValGluProLysGlnGluThr 348
Db 799 GATGAGTTTGATGGTGAATCGCTAGCGAAGCATCACCATTAGCCAGCGTGTGCGTATC 858
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 859 GCCACGCTGGAGCGGACTTCATGACGACTTGAAGGTATCTCTTGACGTCAGGACGCT 918
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
Db 919 GCGGACTTCGAGGCTTTGAGTCGTATAGCCGCGAGACGCTCATCCGACGATCCGC 978
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaThrArgGlySerLeu 408
Db 979 GATCTTGGCAGCGTGGGTGCACTTTTGACGCACTGTACGAGCGCTTACCTCAATGCGTGCTA 1038
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 1039 ATTCTTTGGCAACGCGCGTTCGATCCC-----AACCTTCGTTCCAGCAGAG 1092
Qy 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
Db 1093 GACAAGCTCGATAACACGAGGACGCTGTTGTGAATTCGATCCGACACGCTGCTGAGTCTG 1152
Qy 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIleTyrGlnValHis 464
Db 1153 GTGAGTGAAGTGGCTACGCGCGCTTGAAGCGCGTACGGTACCAAGAGTTTAACTTCAT 1212

```

```
Qy 465 ArgPheAlaArgProGluAlaLeuGlyGly---ThrLeuHisAenThrIleAla----- 481
Db 1213 CGTCCGCTGCGCCCTGAGGCTACCGCGTGTCTGATTATCGTTAAACAAATCGCACCGCAG 1272
Qy 482 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 499
Db 1273 AAGGCGGAGACATTTCCCTGAGGTGATCTTCTGTGTGAGAGCTTGGAGATATCTTG 1332
Qy 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
Db 1333 GAGAAAGCTGAATAGCAATAGGACAGACATAGCTGACGCGAGATCTTGACCCCTGAT 1392
Qy 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534
Db 1393 CCTTCATTCCTGTTGCCGATGCGATTCCGCCGAGGCGAGCCCATTCCTCCGTCCTACCGA 1452
Qy 535 SerGlyHisAlaThrGlnAenGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 554
Db 1453 AGCGGCCACGCTGTGGTGTGCGGCATGTGTGACGATCTTGAAGGCGTTC-----TTC 1506
Qy 555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAsp-----AspGly 571
Db 1507 GACTCCGCGATCGAGATC-----GATCAGGTGTTCCGAGTCTCGACAAAGATGAGGACAAG 1560
Qy 572 LeuGluLeuLeuAenPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeu 591
Db 1561 CTTGTGAAGTCTGCTTTCAAGGGAAT---CTCAGTGTGCGCGTGAATTGAACAAGCTC 1617
Qy 592 AlaValAenValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle 611
Db 1618 GCCGACAAATATTCGATCGCGCGGTAACATGCGAGGTGTCTACTACTTCTCTGACCACTTC 1677
Qy 612 GlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631
Db 1678 GAGTCACTTCTGCTCGTGAGCAGGTTGCGATTGGAATCTTGGAGAGCAAGTCTGACG 1737
Qy 632 PheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1738 TATGGCAGAACTTCTTCTTCAACTTCCGAACTTGATGGAATCTACAATCCAGATC 1794

RESULT 5
AAV56021
ID AAV56021 standard; cDNA; 1791 BP.
XX AC
XX AAV56021;
XX
DT 09-DEC-1998 (first entry)
XX
DE Haloperoxidase, CP.BP02 encoding cDNA.
XX
KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
KW bromine; iodine; ds.
XX
OS Corallina pilulifera.
XX
XX Key Location/Qualifiers
XX CDS /*tag= a
XX FT /product= "Haloperoxidase, CP.BP02"
XX FT /note= "the stop codon is not indicated"
XX
XX JP10248581-A.
XX
XX 22-SEP-1998.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX WPI; 1998-560733/48.
XX
DR P-PSDB; AAW80551.
```

```
XX New haloperoxidase gene - and corresponding vector, transformed host cell
PT and method of preparation.
XX
XX Claim 2; Page 15-18; 20pp; Japanese.
XX
CC This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
CC encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
CC pilulifera. A host cell transformed with a vector containing the
CC haloperoxidase genes can be used for the recombinant production of the
CC enzyme. The haloperoxidases of the invention are vanadium-dependent and
CC have halide specificity of bromine and iodine
XX
SQ Sequence 1791 BP; 390 A; 459 C; 507 G; 435 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,53e-48 Length: 1791
Score: 614.50 Matches: 193
Percent Similarity: 43.04% Conservative: 70
Best Local Similarity: 31.59% Mismatch: 223
Query Match: 17.42% Indels: 125
DB: 2 Gaps: 21

US-10-691-383-2 (1-676) x AAV56021 (1-1791)
Qy 141 ArgAspAsnValAlaPheAlaSerArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGCGCCAAAGGCTTCATTCGATACGCGTGTAGTCGCGCGAGCTGGCACTC---GCCGT 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
Db 88 GGAGTGTGTGCATCGTTTGCATAATGGAGAGGAGCTTCTTACCCCAACTGTGACCCGGC 147
Qy 171 ---ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln--- 188
Db 148 GATCCAAAGCTTCATGTAGCTTCACAAAGGCTTCCACATGACGACATGGCGCTATT 207
Qy 189 ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIle 208
Db 208 ATTAGTCCCGACGATTTCTTGGCCTTCGTCGATCAATAGTCGAGATGAGAAGGAG 267
Qy 209 CysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 ATCCCGACCTTACATTTGGGGCCAGCTCGGACCTCGAGACTGGCTTACCNAATCTGGCGC 327
Qy 222 -----ValAsnProThrAlaAla 227
Db 328 TCAGATCTGGGAATTCCTCTCGAGCTCGAAGTGGCAGGATGGGAAACAGCTCTGCCGT 387
Qy 228 PheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProValProThr 247
Db 388 CTTACCTTCGACCTGGAGGCGCCGACGCGCAGTCGGTTGCCATGCCACCGCGCTGTG 447
Qy 248 LeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArg 267
Db 448 CTCATGAGCCCTGAGCTCATCGCGAGATGCGCAACTGTACCTGTGGCGCTGGACGC 507
Qy 268 AspValProPheMetGlnTyrGlyThrAspGlu-----IleThrThrAlaAla 284
Db 508 GACATCGAGTTTACGAGTTTCGAGTCCCAAGAACCGCAGCTTTTATTAGTCTGCTATT 567
Qy 285 AlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAla----- 298
Db 568 GAACGGCTTAACGGCTGGAGTGGTTC---AACACACCGGCAAGCTTGGTGATCCGCT 624
Qy 299 -----ValSerIleGlySerAspGlyThrValAspPro 309
Db 625 GCGGAATTCGCTCGCCGCTCGCGGTGAGTGCACCTCGGAAC----- 666
Qy 310 PheSerGlnLeuPheArgAlaThrPheValGlyValGluThrPheValSerGln 329
Db 667 -----TTGTTCCCGCGGTATTCTTCCCGGTTCCTGAGGTCCGCCCGCTATCTCAGCCAG 717
```


Qy	23	uThrSerGluGlnArgAlaIuAluysGlyTTrpArgArgGlnLeuGluGlyGluIuysSerIeuGl	43
Db	6786	CACCGCGGGCTCGACGCGCGCGGTACGCGCGAGCGTCAAGGGGTACACGAACCTCGA	6845
Qy	43	yPheHis-----ProSer-----GluThrProTyrlleLy	53
Db	6846	CGCCACCGGGTGTCTGTACGGGTGCGGAAACCGGCATTGTGCGAGCAGCGCAAGCA	6905
Qy	53	sTyzLeuGluGlySerGluThrTrpLysValysLeuProThrAspGlyIleSer--	72
Db	6906	GCAGTGCAT---CTCCGGCCTGGCGGCTCTGGCGTG---ACCAGGGTGTTCAGCA	6959
Qy	73	-----AlaSerLysIleLeuGlyIleMe	81
Db	6960	CGCGTGGGCGGGCGGACGCGCGCTGTGTGGGTCCGATCCGTGAGGCGAGCAGTGT	7019
Qy	81	tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh	101
Db	7020	GGAGCGCACCGG-----GGATCGTCTGGCGGG-----	7050
Qy	101	eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly-----	118
Db	7051	-----ACGTCCGACGCGCTCTTC-----CATCGGGCTTCTCTCGA	7085
Qy	119	---GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs	134
Db	7086	CGAGGACGGCGCGTGTGGCGGGTCGAGGACGCCGAGTTACCGCGGGCGACCTTGG	7145
Qy	134	pProThrAlaProAsnArgArgAspAsnValAlaPheAla-----	147
Db	7146	GCCGGCACTCCCGGTGAGGCCGACGCGCATCTGTGGCATGTCGCGAGCGCTCGCGTCC	7205
Qy	148	-----SerArgArgAspAlaAlaArgArgG	156
Db	7206	GGTCTCGAGACGCCGGTTGGTACGGGGAGTGCACAGTCGAGAGCCGTGCGCGCGA	7265
Qy	156	uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly-----	169
Db	7266	GGCGAGCGCTCGGTGACCGTGTGGCGGTCGCGAGCGGGCGCGCGCTCTGTCGGGTC	7325
Qy	170	-----GluThrAspLeuAlaTh	175
Db	7326	GCTCGCGGAGACCGTCGCCGACCTGTCTCAAGTTCGACCTGCGGACATCGACCTCGAC	7385
Qy	175	rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh	194
Db	7386	GCACCTTCAGCGTACGGCTTCGAGTCCATCGCGCTGGCCAAACTGCGCTCGGAACCTCA	7445
Qy	194	eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl	214
Db	7446	CGGGTCTCTCGGACGGACCTCACCCCGCGCTCTTCTTCAGTGTCTCCGAC-----	7497
Qy	214	aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGl	234
Db	7498	-----ATCCGACGCGCTCGCGAGTACCTGTCTCGACCGCTACGG	7535
Qy	234	yProAlaPheSerAlaThrThr-----IleProValProThrIleuSe	249
Db	7536	CCCCGAGTGAAGCTCCCGACGAGCGGACGCCCGCGCGGTGCGCCGCCACCGCGCC	7595
Qy	249	rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspVa	269
Db	7596	GTCCCCA-----GT	7604
Qy	269	lProPheMetGlnTyrglyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGl	289
Db	7605	GCCGATCCGCGACCGCGCGCGGACGAC-----GACGCGGTGGCCATCTGCGCGC	7655
Qy	289	yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr	309
Db	7656	TGCCGGAAGCTTCCCGCGCGGAC-----GACCTGGACAC	7691

Qy	309	oPheSerGlnLeuPheArg	-----AlaThrPheValGly	-----Va 321
Db	7692	CTTTCGACGAGCTGCGCGCGGCGAGGACCTGATCGCCGACATACCCCGCGCACCGCTT 7751		
Qy	321	lGlUThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleTh 341		
Db	7752	CGACGGGGGGCCCTTACCGGAGGTGCTGCGCGGGCGGACTTCCGAAGTTTCCCGCGCG 7811		
Qy	341	rValGluProIysGlnLeuThrPheAlaProAspLeuAsnTyrMetValAspPheAspG1 361		
Db	7812	GATCGAG---GGCGTGAGCCGCTTCGACGCGGACTCTTCCACCTGTCGGGTGGAGGC 7868		
Qy	361	u-----	-----TrpLeuAsnIleGlnAs 367	
Db	7869	GGAGCTGATGGACCCCGCAGCACCGGCTGGCCCTGGAGACCGTGTGGCGCGCGCTGGAGAA 7928		
Qy	367	nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378		
Db	7929	CGGCGGCTACGCCCCCGCGGCGCTTCCCGCGAGAACACCGGGGTCTACTTGGCGCTCTCCGG 7988		
Qy	378	pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391		
Db	7989	CAGCGACTACCACCACTGTCTCAACGCCAGTGGCGTGGCACCCGACCGGCTTCACCGCCAC 8048		
Qy	392	-----ArgValSerPheValAspAsnIleAsn----- 400		
Db	8049	CGGCAACGCCCACTCGATGTGCGCAACCGATCTCTACGTCTCGACGTGACCGGGCC 8108		
Qy	401	-----	-----ThrG1 402	
Db	8109	GAGCGAAACCGTCGACACCGGCTGCTCCAGTCTGCTGCGCTGCACCGCGCCGTGCA 8168		
Qy	402	uAlaTyrArg-----GlySerLeuIleLeuLeuG1 412		
Db	8169	GCATATCCGTGCGGCGGATGCGAGATGGCCATCGCGGGCGGTGTCAACTGCTGTGTGAG 8228		
Qy	412	uLeuG1AlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432		
Db	8229	CGTGGACACCTTCGCGCGACGACATATGGCGGC---ATGCTCAGCCCCGACGGCGCGCTG 8285		
Qy	432	aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451		
Db	8286	CAAGACCTTCTCGCGCGCGCGACCGCTACGTCCGTCGAGGGCGGTTCGCGCGGTGTCT 8345		
Qy	451	uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPh 466		
Db	8346	GCTCAAGCCCTCGCCAGCGCGACGGGACGGGACGGCGAGCCATCTGGGGGTCTCCGGGG 8405		
Qy	466	eAlaArg----- 468		
Db	8406	CAGCGCGAGAACACGGGCGCGCGCGCTTTCGCTGCTGCGCCCAACGGCAAGCGCA 8465		
Qy	469	-----ProGluAlaLeuGlyGlyThr--- 475		
Db	8466	GGCGCGCCTGATCCAGACGCCATGCGCGGCATCGACCGGACAGCATCGGCTACGTCA 8525		
Qy	476	-LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495		
Db	8526	GGCGCAGCGGCGGACCGGC---CTGGGCGACCCGCTCGAGGTCAACGCCCTCGACAG 8582		
Qy	495	nAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515		
Db	8583	CGCCTACCGCGCCTTCGCCACCGCCGAG----- 8610		
Qy	515	lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535		
Db	8611	-----GGCGGCGCGCGCACCGCGCGCGCGCGCTG 8639		
Qy	535	r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546		
Db	8640	CGGCGCTCGGCTCGGTGAAGAACCAACATCGCGCCACGCGGAGTCGGCGCGGGCGCTGCCCG 8699		
Qy	546	rValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562		

```
Db      8700 AGTGTGAAGGTGCTGTCCCATG---CCTCACCGGAGCTGCCGCGCCTTGGACTG 8756
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      563 -----AsnProValPheProSerAspGlyLeuGluLeulle 575
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      8757 CGACCGGCTCAACCGGACCTGCGCTCGACGGCGGATTTCGAGTCTGTA 8805
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AAD54217
ID      AAD54217 standard; DNA; 52101 BP.
XX
AC      AAD54217;
XX
DT      17-JUN-2003 (first entry)
XX
DE      Streptomyces platensis subspecies rosaceus dorriginocin DNA.
XX
KW      Polyketide biosynthesis; dorriginocin; DORR; lactimidomycin; LACT; gene;
KW      ds.
XX
OS      Streptomyces platensis.
XX
FH      Key Location/Qualifiers
CDS     complement(67..3720)
FT      /*tag= a
FT      /product= "ORF1 protein"
FT      /tag= b
FT      /product= "ORF2 protein"
FT      /note= "No start codon"
FT      /partial
FT      /tag= c
FT      /product= "ORF3 protein"
FT      /tag= d
FT      /product= "ORF4 protein"
FT      /tag= e
FT      /product= "ORF5 protein"
FT      /tag= f
FT      /product= "ORF6 protein"
FT      /note= "No start codon"
FT      /partial
FT      /tag= g
FT      /product= "ORF7 protein"
FT      /note= "No start codon"
FT      /partial
FT      /tag= h
FT      /product= "ORF8 protein"
FT      /tag= i
FT      /product= "ORF9 protein"
FT      /tag= j
FT      /product= "ORF10 protein"
FT
XX      WO200288176-A2.
XX
XX      07-NOV-2002.
XX
XX      26-APR-2002; 2002WO-CA000591.
XX
XX      26-APR-2001; 2001US-0286346P.
XX      (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX      Farnet CM, Zazopoulos E, Staffa A, Yang X;
XX      WPI; 2003-201222/19.
XX
```

```
DR      P-PSDB; AAE35484, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
DR      AAE35490, AAE35491, AAE35492, AAE35493.
XX
XX      Novel isolated or purified polypeptide involved in biosynthesis of
PT      polyketide dorriginocin or polyketide lactimidomycin, useful for preparing
PT      dorriginocin or lactimidomycin.
XX
PS      Claim 1; Page 85-113; 312pp; English.
XX
CC      The invention relates to novel proteins involved in the biosynthesis of
CC      polyketide dorriginocin (DORR) or lactimidomycin (LACT) biosynthesis by
CC      microorganisms. Sequences of the invention allow direct manipulation of
CC      dorriginocin, lactimidomycin and related chemical structures via chemical
CC      engineering of the enzymes involved in the biosynthesis of dorriginocin and
CC      lactimidomycin. They are useful for introducing chemical handles into
CC      normally inert positions that permit subsequent chemical modifications
CC      and facilitate the development of polyketides. The genes and proteins of
CC      the invention can also be used to generate a focused library of analogues
CC      around a polyketide lead candidate to fine-tune the compound for optimal
CC      properties. They are useful for generating antibodies specific for the
CC      polyketide biosynthesis. The present sequence is Streptomyces platensis
CC      subspecies rosaceus DORR DNA
XX
SQ      Sequence 52101 BP; 7364 A; 17894 G; 6730 T; 0 U; 0 Other;
      Alignment Scores:
      Pred. No.:      3.37      Length:      52101
      Score:      131.00      Matches:      152
      Percent Similarity:      30.12%      Conservative:      82
      Best Local Similarity:      19.56%      Mismatches:      254
      Query Match:      3.71%      Indels:      289
      DB:      Gaps:      38

US-10-691-383-2 (1-676) x AAD54217 (1-52101)
Qy      5 AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      14734 GCCGAGATGAGCAGCGCGCCGCCGACCTCGACGGGACGACTTCTACGCGGCTCGG 14793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      23 uThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGluGlyGlyLysSerLeuG 43
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      14794 CACGCGGGGCTCGACGCGCGCGCTACGCGCGCAGCGTCCGAGGGGTCCACGAACTCGA 14853
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      43 yPheHis-----ProSer-----GluThrProTyrIleLe 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      14854 CGCCACCGGCTGCTGTACGGGTGCGCGGACCGGCATGTGCAGACCGGCACAGCA 14913
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      53 sTyrLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      14914 GCACGTGCAT---CTCCGCGCTGGGCGCTCGTGGGCTG---ACCCAGGGGTGTTCAGCA 14967
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      73 -----AlaSerLysIleLeuGlyLysIleMe 81
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      14968 CGCGTGGGCGCGGCGCGACCGCCGCTGGTGGCGGTGCGATTCGTCAGGCGGAGAGTG 15027
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      81 tAlaArgValArgIleAlaThrAlaLeuAlaValLeuAlaAlaProCysLeuAlaPh 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      15028 GGAGCGCACCCCGG-----GGATCTGCTGGCGCGG----- 15058
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      101 eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      15059 -----ACGTCCGACGCGCTCTTC-----CATCGCGCTTTCCTCGA 15093
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      119 -----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      15094 CGAGGACGCGCGCTGCTGGGCGGCGGTGCGAGCGCGGATTCACCGCGGCGCACCTGGA 15153
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      134 pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      15154 GCCGGCACTCCCGGTCAGCGCGGACCGGCACTCGTGGCACTGCCGCGCGCTCGCGTCC 15213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      148 -----SerArgAspAlaAlaArgArgL 156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 15214 GGTCTGGAGACCGCGGTGGTGTACGGGGAGTGCAGAGTCGAGGCGGTGCGGCCGGA 15273
Qy 156 uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169
Db 15274 GCGCAGCGCGTCCGTGACCGTGGCGGGTTCGCGGACGGCGCGCGCGTCTGTCGGGTC 15333
Qy 170 -----GluThrAspLeuAlaTh 175
Db 15334 GCTCGCGGAGACCGTCCGCGACCTGCTCAAGTTTCGACCTGGCGGACATCGACCTCGACAC 15393
Qy 175 rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspPhe 194
Db 15394 GCACTTCACCGCTACCGCTTCGAGTCCATCGCGCTGGCCAACTCGCCTCGGAACCTCAA 15453
Qy 194 eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214
Db 15454 CGGCGTCTCGGCACGACCTCACCCCGCGCGTCTTCTCGAGTGCTCCGAC----- 15505
Qy 214 aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGl 234
Db 15506 -----ATCCGCGCTCGCGGAGTACCTGCTCGACCGCTACGG 15543
Qy 234 yProAlaPheSerAlaThrThr-----IleProValProThrLeuSe 249
Db 15544 CCGCGAGCTGAGCTCCCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15603
Qy 249 rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspVa 269
Db 15604 GTCCCA-----GT 15612
Qy 269 lProPheMetGlnTyrGlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGl 289
Db 15613 GCCGATCCCGCGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15663
Qy 289 yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309
Db 15664 TGCGGACCGTTCCTCCCGCGCGCGAC-----GACTTGGACAC 15699
Qy 309 oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321
Db 15700 CTTCGCGACAGCTGCGCGCGCGCGGAGGAGCTGATCGCGGACTACCCCGCGCGCGCTT 15759
Qy 321 lGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIleTh 341
Db 15760 CGACGGGGCGCTTACGCGAGGTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 15819
Qy 341 rValGluProLysGlnThrPheAlaProAspLeuAsnTyrMetValAspPheAspGl 361
Db 15820 GATCGAG---GGCGTGGACCGCTTCGACGCGGACTTCTTCACCTGTCGCGGTGGAGGC 15876
Qy 361 u-----TrpLeuAsnIleGlnAs 367
Db 15877 GGAGCTGATGACCCCGCAGCACCGGTGGCGCTGGAGACCGTGTGGCGCGCTGGAGAA 15936
Qy 367 nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378
Db 15937 CGCGCGCTACCGCGCGCGCGCTCCCGGAAACACCGGGGTCTACTTCGCGGTCTCCG 15996
Qy 378 pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391
Db 15997 CAGCGACTACCAACCTGCTCAACCGGACGTGGCGTGGCACCCGAGCGCTTTCACCGCCAC 16056
Qy 392 -----ArgValSerPheValAspAsnIleAsn----- 400
Db 16057 CGGCAACGCCCACTCGATGTGGCCAAACCGGATCTCTAGCTGTGAGCTGACGGGCC 16116
Qy 401 -----ThrGl 402
Db 16117 GAGGAAACCGTCGACACGGCTGCTCCAGTCTGCTCGCGTGCACCGCGCGCTCGA 16176
Qy 402 uAlaTyrArg-----GlySerLeuIleLeuGlu 412
Db 16177 GCACATCCGTCGCGCGCGGTGCGGATGCGCATCGCGGGCGGTGTCAACCTGCTGTGAG 16236

Qy 412 uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432
Db 16237 CGTGGACACCTTCGCGCGGACACATGGCGGC---ATGCTCAGCCCGACGCGCGCTG 16293
Qy 432 aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451
Db 16294 CAAGACCTTCGCGCGCGCGCGGACCGCTACGCTCCGCTCCGAGGCGCTGCGCGGTGCT 16353
Qy 451 uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPhe 466
Db 16354 GCTCAACCGCTGCCCGAGCGGAGCGGAGCGGACGCCATCTGGGGCGCTGCTCCGGGG 16413
Qy 466 eAlaArg----- 468
Db 16414 CAGCGCGAGAACACCGCGCGCGCGGTTCGCTGACCGCCCAACGCAAGGCGCA 16473
Qy 469 -----ProGluAlaLeuGlyIleThr-- 475
Db 16474 GGCGCGCTGATCCAGGACGCCATGCGCGGATCGACCCGACACATCGGCTACGTGCA 16533
Qy 476 -LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495
Db 16534 GGCGCAGCGGACCGCGCACCGGC---CTGGGAGACCGGTGAGGTCAACGCCCTCGACAG 16590
Qy 495 nAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVa 515
Db 16591 CGCTACCGCGCTCGCGCACCGCGAG----- 16618
Qy 515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535
Db 16619 -----GGCGGCGCGCGCGACCGCGCGCGCGCGGTG 16647
Qy 535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
Db 16648 CGCGCTCGGTCTGAGAGAACACATCGCGGACCGGAGTGGCGCGCGCGCTGGCGCG 16707
Qy 546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562
Db 16708 AGTGCTGAGGTGCTGCTGCCATG---CGTCCCGCGAGTGGCGCGCGCGCTTGCACGT 16764
Qy 563 -----AsnProValPheProSerAspAspGlyLeuGluLeuLe 575
Db 16765 CGACCGGCTCAACCCGCGCTGCGCTCGACGCGGATTCGAGGTCTGA 16813

RESULT 8
ADA30090
ID ADA30090 standard; DNA; 11679 BP.
XX
AC ADA30090;
XX
DT 20-NOV-2003 (first entry)
XX
DE DNA encoding Acinetobacter baumannii protein #1377.
XX
KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX vaccine; plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
PN US652958-B1.
XX
PD 13-MAY-2003.
XX
PF 04-JUN-1999; 99US-00328352.
XX
PR 09-JUN-1998; 98US-0088701P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Breton G, Bush D;
XX
DR WPI; 2003-576092/54.


```
QY 520 ----ProGlnAla-----Ile 523
PT ||| |||
DB 2341 GCACCTGATGACCAAGTAATTAACCCGTTAAACGGGACAGACCGGATTACAGGTACGGCA 2400

QY 524 GlnValGlySerProThrHisProSerTyProSerGlyHisAlaThrGlnAsnGlyAla 543
DB :::: ||||| |||
DB 2401 GAGCCTGGTTCAACAGPACTGTGACTTATCCAGATGCGAGTACAACA----- 2448

QY 544 PheAlaThrValLeuLysAlaLeuLeGlyLeuAspArgGlyGlyGluCysPheProAsn 563
DB :::: ||||| |||
DB 2449 -----ACAGTTGTTGACGAGCCGGAT---GGCAGTTGGACAGTACCAAC 2490

QY 564 ProValPheProSerAspAspGlyLeuGluLeuLeuLeuPheGluGlyAlaCysLeuThr 583
DB ||| |||
DB 2491 CCAGGTTAAATGATGCGCAT----- 2511

QY 584 TyrGluGlyGluLeuLeuLeuLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGly 603
DB :::: ||||| |||
DB 2512 -----AAAGTTACAGCAATGCTACAGATCCAGCA---GGCAACCCATCATACCA 2559

QY 604 IleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArg 623
DB ||| |||
DB 2560 GGTACAGCTACTGTTGATGAGTGGTCCAAATACCATGGTGTATTTACAGTTGAT 2619

QY 624 ThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArgLeuPhe 643
DB :::: ||||| |||
DB 2620 TCAGTAACAGCTGACAATGTGATTAATGCATCAGAACATCAGCAACGTTACTGT--- 2676

QY 644 ThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCys 663
DB :::: ||||| |||
DB 2677 ACAGGTGATATGAAAAAGTTCCGGCAGATGCGCAAAATACAGTGGTCACTGTTGTGATC 2736

QY 664 SerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
DB :::: ||||| |||
DB 2737 AATGCCAGAGCTATACTGCAACTGTAGATAGCACGCA 2775

RESULT 9
ACA43739
ID ACA43739 standard; DNA; 18930 BP.
AC ACA43739;
XX
XX
DT 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #25396.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Pseudomonas putida.
XX
XX W0200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 08-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JB, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX P-PSDB; ABU9869.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
```

```
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 31609; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
```

Sequence 18930 BP; 3570 A; 6691 C; 5875 G; 2794 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.79	Length:	18930
Score:	124.00	Matches:	147
Percent Similarity:	30.99%	Conservative:	56
Best Local Similarity:	22.44%	Mismatches:	226
Query Match:	3.51%	Indels:	229
DB:	7	Gaps:	31

US-10-691-383-2 (1-676) x ACA43739 (1-18930)

QY	111	ProGluGluHisLysHisThrGlyGlyArgHisLeuGln-ThrCysThrAsn----	128
DB	573	CCTGATCGGCAGCGGCACAGTGGTCCGACGCAACTTCAACGTGACGTTGAATGCGCC	632
QY	129	-----SerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh	146
DB	633	GCAGATCAACACGCGAAAAACCTTGATGTCACTGACGCGCCGCGGTATGTCTCG--	690
QY	146	eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI	166
DB	691	-----GCGCCTGGCGCTGTCTACTGCGACGCGACCACTGCACCATCTG--	732
QY	166	eThrAsnGlyGluThrAspLeuAla-----ThrMetPheHisLysSe	180
DB	733	-----GCGCGCAGCCGACCTGCTATCAATGAGAGGGGTAACACCCCTTACCGGTCTGTC	785
QY	180	rLeuProHisAspGlu-----LeuGlyGlnValTh	190
DB	786	CGAGCCCGGCTCGACGCGTTTCCGTGCGAGCGCAGGTGGCGGTATTGTGCGCACTGGCGT	845
QY	190	rAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG	210
DB	846	GGCCCGGT-----GCGGACGCGCAGTTACGATTACCT	878

Qy	210	uAepValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaI	230
Db	879	GCAACCGCCACAAAGCGAC-----GGCCAGGCGCTGGAGATCAGCGCCGACAGCGCGC	932
Qy	230	eAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal-----ProThrLe	248
Db	933	TGGCAATATCTCGCCAGTTCCAGCATCAGCGGG--CCGGAGCGTCGACAACCCAGATAC	989
Qy	248	uSerSerProGluLeuAlaGlnLeuAlaGluLeuTyTirMetAlaLeuAlaArgAs	268
Db	990	CACAGCCCTGAAACACCGACCGACCTG-----GCCCTGGCGGATGG	1031
Qy	268	pValProPheMetGlnTyGlyThrAspGluIleThrThr-----ThrAlaAlaAl	285
Db	1032	CGTCACATTTTCCCGGACGCGGTGAGCTGTGTCTACCGTCAGGTGCGCATGCTGCGCG	1091
Qy	285	aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspG	305
Db	1092	TAAACCTGATTGGCACCGGT-----GTGTCGTGGTCCGATGG	1127
Qy	305	yThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr	325
Db	1128	C-----CTGTTACGCTCACCTCTGCTGCTGCCAGACC-----	1161
Qy	325	oPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy	345
Db	1162	-----AATGTTGAAGCACTCGAGCTGCTGCTGTGGATGCCGCG	1199
Qy	345	sGlnGluThrPheAlaPro-----AspLeuAsnTyTirMetValAspPheAspGluTr	362
Db	1200	CGGCAACAGAGCTCTGCCCATTTGCAATTTCGATCGCGCGGACATCACACCGCTGACGCAGT	1259
Qy	362	pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr	382
Db	1260	AACCAATATTACAGTTGGTGTCTGACGG-CTTGGCCCTGACGCGTCTGGCGGACGCCAGTG	1318
Qy	382	gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs	397
Db	1319	CCACTGTGGAAGTGGCGATGCCAATGGCACTGTGATCG-GCAGGGTGTGTTGGGTGCC	1377
Qy	397	pAsnIleAsnThrGluAlaTyArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSe	417
Db	1378	-AAC-----GGCACCTCTCTGATCGACCTGCCCTCCCGACGCCGCGC	1412
Qy	417	rArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPh	437
Db	1413	ACAGCGGGCGAGCAGTTGACCTGTGTGTGCAGACCGAC---CCAGCGGTAAATGCTTCAGT	1469
Qy	437	eGlyThrSerHisTyPheArgLeuIleGlyAlaAlaGlu-----	450
Db	1470	GGCTACCGAGTACGATGTACCACTGACCAACCGCGCGCGACAGCCGAGCAATCTGGCCAT	1529
Qy	451	-----LeuAlaGlnArgAlaSerCysTyTyrGlnLysTrpGlnValHi	464
Db	1530	CGATGCTGATGGCACCAACCTCTCACCGGTACCGCGCGCGCTGGCAGCGCTGTGGAGTGCA	1589
Qy	464	sArgPheAlaArgProGluAlaLeuGlyThrLeuHiAsnThrIleAlaGlyAspLe	484
Db	1590	T-----GACGCCAACGGCAGCGCTGATTGGCATGCCAATGTGCCAACGC	1631
Qy	484	uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuLysAr	501
Db	1632	CGATGGCAGCTTCAGCATCGAATCGCGCACAGCAACCGCGCAACTGCTGGACGT	1691
Qy	501	gValAla-----	503
Db	1692	GGTTGCCATCGATGACGGCGCGTTCCTCGCTGCCCGCACAGATCACCGCGCCCGACAT	1751
Qy	504	-----GluIleAsnAl	507
Db	1752	CACGCGCCCTCGCGCACCGACCGAATGTCGGGTGAGCGCGGACGGCAGCGTTATTCACCGG	1811
Qy	507	aAlaGlnAsnProAsnAsnGluValThrTyLeuLeuProGlnAlaIleGlnValGlySe	527

Db	1812	CCGTGCCGAACCCGGGCGAGCACCGGTGGCATCTGGTGGTCCGACGGAACCACTGGGCAC	1871
Qy	527	r-----ProThr-----HisProSerTyrProSerG1	536
Db	1872	TGCGGTGGTCCGGCCGACCGCGGTTCAGCTCAACCTCAACCCGCCACAGTTGACGG	1931
Qy	536	yHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspAr	556
Db	1932	CGAAGTCTGCAA-----GCCACGGCGACC-----GATGC	1961
Qy	556	gGlyGlyGluCysPheProAsnProValPhe-----ProSerAspAspGlyLeuGlu--	573
Db	1962	CGTGGCAACACATCGCCAAACAGCAGCGCAGTTACCGCACCGGATATCGACGGGTTCGATAC	2021
Qy	574	-----LeuIleAsnPheGluGlyAlaCysLeuThrTy	584
Db	2022	CACGCCACCGGACGCCCGAGTACTGTGGTGGATCGGCTCGCGGACGACCACTCAGTGG	2081
Qy	584	rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyI1	604
Db	2082	TCGAGGCGAA-----	2091
Qy	604	eHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgTh	624
Db	2092	-----GCCGGCTCTACCGTGCGAGGTACGTGA	2117
Qy	624	rLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArgLeuPheTh	644
Db	2118	CGCGGCGAGGCAACATC-----CTCGGCAC	2141
Qy	644	rGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIle	658
Db	2142	CGGCACCGTCGCC-----GGGATGCACCTTCACTGTC	2175
RESULT 10			
AAZ55887			
ID	AAZ55887	standard; DNA; 68750 BP.	
XX	AAZ55887;		
AC			
XX			
DT	15-SEP-2003	(revised)	
DT	10-APR-2000	(first entry)	
XX			
DE	Sorangium cellulosum	68.75 kb contig.	
XX			
KW	Epithilone biosynthesis; type I polyketide synthase; taxol substitute;		
KW	anticancer; ds.		
OS	Polyangium cellulosum.		
Key	Location/Qualifiers		
FT	1..1826		
FT	/*tag= a		
FT	/partial		
FT	/product= "Partial Orf 1 protein (AAZ55880)"		
FT	/note= "No initiation codon given in the specification"		
FT	complement(1900..3171)		
FT	/*tag= b		
FT	/product= "Orf 2 protein (AAZ55881)"		
FT	3415..5556		
FT	/*tag= c		
FT	/product= "Orf 3 protein (AAZ55882)"		
FT	complement(5612..5992)		
FT	/*tag= d		
FT	/product= "Orf 4 protein (AAZ55883)"		
FT	6226..6675		
FT	/*tag= e		
FT	/product= "Orf 5 protein (AAZ55884)"		
FT	7610..11875		
FT	/*tag= f		
FT	/product= "Type I polyketide synthase, EPOS A (AAZ558873)"		
FT	11872..116104		

PF 21-MAY-2002; 2002CA-02387401.
 XX
 PR 21-MAY-2001; 2001US-0291959P.
 PR 03-DEC-2001; 2001US-0334604P.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 XX
 XX Staffa A, Farnet CM, Zazopoulos E;
 PI
 XX WPI; 2002-692206/75.
 DR P-PSDB; ABG97437.
 DR
 XX
 PT New enediyne polyketide synthase polypeptides involved in synthesis of
 PT the warhead structure in enediyne compounds useful in design of new
 PT enediyne compounds and identification of new enediyne biosynthetic loci.
 XX
 PS Claim 1; Page 183-186; 310pp; English.
 XX
 CC The invention relates to purified/isolated polypeptides involved in
 CC synthesis of the warhead structure in enediyne compounds, or their
 CC fragments or sequences 75% homologous to them. The polypeptides comprise
 CC five families of proteins: PKSE (polyketide synthase), TEBC (thioesterase
 CC unique to enediyne biosynthetic loci), UNBL, UNBV and UNBU (unique to
 CC enediyne biosynthetic loci L, B and V). The proteins are isolated from 10
 CC different micro-organisms. Also included are the nucleic acids encoding
 CC the proteins, antibodies specifically binding the polypeptides, and a
 CC computer system/computer readable medium comprising the sequences
 CC referred to above. The polypeptides and polynucleotides are useful in
 CC genetic engineering to design new enediyne compounds (which are potential
 CC anti-cancer drugs) and identify new enediyne biosynthetic loci. They can
 CC also be used to identify enediyne biosynthesis genes/gene fragments from
 CC samples containing genomic DNA e.g. in biomass from an environmental
 CC source (such as a mixed microbial culture/population of organisms), pure
 CC cultures or genomic libraries obtained from a mixed population of
 CC organisms or from pure culture. The present sequence encodes an enzyme of
 CC the enediyne warhead structure biosynthetic pathway
 XX
 SQ Sequence 5820 BP; 846 A; 2105 C; 2071 G; 798 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.92 Length: 5820
 Score: 123.00 Matches: 146
 Percent Similarity: 31.73% Conservative: 59
 Best Local Similarity: 22.60% Mismatches: 249
 Query Match: 3.49% Indels: 192
 DB: 6 Gaps: 32

US-10-691-383-2 (1-676) x ABS78676 (1-5820)
 Qy 4 HisAlaAlaPheThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeu 23
 Db 1816 CACCGGACGACATCTACCGCGCGCGCAACCTG---TCCACCGCGCGCGACAGGTGCC 1872
 Qy 24 ThrSerGluGlnArgAlaLys-----GlyTrpArgGlnLeuGluGlyLys 40
 Db 1873 ACCGACGTCCGCCAGCGCGCATCGTCACCGTTCGTCGCCGGCTCGGGTCTGAAG 1932
 Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
 Db 1933 AGCCTCGGCATCGAGGGCGGCGCCGTC-----ACGGGCGCACACGCTTCGGCGAGCTCACC 1986
 Qy 61 -----TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeu 77
 Db 1987 GCCCTGCACCTGGGCGCGGCTCACCAGCGCGAAGTGTCTAACTGGCCAGATCCGC 2046
 Qy 78 GlyLysIleMetAlaArgVal----- 84
 Db 2047 GGCAAGGTGTGGCGACCGCAGCGAGCGGGGCCCATGGCGGCGATCGGGCCACG 2106
 Qy 85 ----ArgIleAlaThrAlaLeuAla----- 91
 Db 2107 CCGAGTGTCCCGAGGCGCTGGCGCGAGGCGGCGAAGAGGTCTGTCATCGCGGGCTACAACGCC 2166

92 -----ValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAla----- 106
 2167 CCGGACGAGACTGTCTTCCGACCGCGGAGCGGATCGTCCGCGTGGTGGCGCGTCC 2226
 107 -----SerGly 108
 2227 CGCGCGGAGGGGTCAACCGCGCGCGCATCAAGTCTCGCACGCTTCCACTCGCCCGCG 2286
 109 ValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln----- 124
 2287 GTCTGCCCGCGCGCGAGCGATGACCGGGGAACTCGCCGCGATCGACTTCGGCGCGCTC 2346
 125 -----ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
 2347 GACGGCGCGTCTCTCCACGCGTGACGCGTGACGCTCTGCACCGCGCGGAGACCTCGCG 2406
 141 -----ArgAspAsnValAlaPheAlaSerArg---ArgAspAlaAlaArgArgGlu 156
 2407 GATCTGTCTGCGCGACCGAGGTGCTTCCGCGTTCGCTCGTGGCGCGCGGAGGTC 2466
 157 ArgAspGlyThrGlyThrValCysGlnIleThrAsnGlyGlu----- 170
 2467 GCCGAGCGCAGCACCTGTGTGATCGAGGTGCGCGCGCGCGCGTCTCACCGCGCTGCTC 2526
 171 -----ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAsp 184
 2527 GGCACCATCGCGCGCGGTACCCCGGTGCTTTCGATCGACACCGACGCTGACGCTCGCG 2586
 185 GluLeuGlyGlnValThrAlaAspAspPheAlaIle-----LeuGluAspCysIle 201
 2587 CCGGTGTGAAGGTGCGCGGTTCGCGTTCGCGCGCGCGAGCTGGAGAGCTCCACG 2646
 202 Leu-----AsnGlyAspPheSerIle----- 208
 2647 CTGTTCGACGCGCGGTGCTGCGCGCGCTCGCGCGAGCTGCTGCTCTCTCGCC 2706
 209 -----CysGluAspValPro-----AlaGlyAspProAlaGlyArgLeuValAsn 223
 2707 AGCCCGTGCAGGCGCGCGTCCATCGCGCGCTGCTGTGACCGCGCGAGTCCGCGAG 2766
 224 ProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIlePro 243
 2767 CCGCGCGAGCGCGCGCGGACCGCTCGGAAAGCGGTGGCGAGCAGCAC- 2817
 244 ProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMet 263
 2818 ---CTGACCTCTGCGC-----AAGCTCGCTCGGAACGGTCTGAGTCTGCGCTCGAA 2868
 264 AlaLeuAlaArgAspVal---ProPheMetGlnTyrGlyThrAspGluIleThr----- 280
 2869 CGCGTCAACCGCGCACCGCATCCCTCGACGATCGCACCTTTCGTCTGATCAGCTCGCG 2928
 281 -----ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsn 295
 2929 CAGCTGTCAACGACGTGACCGAGGGCGCTCGCGCGCGCGCGTGGAGGGTATGCCAAC 2988
 296 LeuAspAlaValSerIleGlySer---AspGlyThrValAspProPheSerGlnLeuPhe 314
 2989 TTCGCGACCGTGTCTCTGCGCGAACTCGCGAGATGATCGACGCTCGCGCGAGACCGCC 3048
 315 Arg-----AlaThrPheValGlyValGluThrGlyProPheValSer 328
 3049 AAACCGCGCGACACCAACGAGGCGCGGTCTCGCGCGT-----GGCGCGTGG----- 3096
 329 GlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
 3097 -----GTCCGGCGGTTCGCGGTGGAGTACGTC---GTCCGCGCGGAGCGCG- 3138
 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
 3139 ---TCGCGCGCATCTCGCGACGGGTATCTCCACCGCGGAGTGGAGCGGTTCGCGCGCGCC 3195
 369 GlyProAlaGlyProGluLeuLeuAspGluLeuArgPheIleArgAsnAlaArg 388


```
Db 3196 GGTCAACCA-----CTGCGGAGCGCTGCGC-----GCG 3225
Qy 389 AspleuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTrpArgGlySerLeu 408
Db 3226 GCATTGGCCAGCGGGCGTGGCGACGGT----- 3255
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 3256 GTCCTCTCTGCTGAAAGCCGACAGACGCTTCCGGCGACGTCGGCGTCTTCTCGACGCG 3315
Qy 429 AspArg-----GlnAlaGlyPhe 434
Db 3316 GCGCGCGCGGTGTCGGCGCGCCCAACGGCAGCGGTCGTCGTGGTGCACGCGTCTC 3375
Qy 435 ValAsnPheGlyThrSerHisTrpPheArgLeu----- 445
Db 3376 GCGCGCTCGGGCTGGCGAGACGCTCGGGTGGAGGACCGTCGGCGCGCCACACGATC 3435
Qy 446 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTrpGlnLysTrpGlnValHisArg 465
Db 3436 GTCGACCTCGCGACCTCGGCGCG----- 3459
Qy 466 PheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAsp 485
Db 3460 ---GTGACCCCGAGCCCTCGACCGCGCGTG---TCCACTGTGGTACCGNAGTGGCG 3513
Qy 486 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIle 505
Db 3514 GCGACGACCGACTTCAGCGAGTCCGTACGACACCGCGGAGTCCGACGCGTCCCGNAG 3573
Qy 506 AsnAlaAlaGlnAsnPro 511
Db 3574 CTCGCGCGCTGACCCCG 3591
```

RESULT 13

ACF67367_38

Continuation (39 of 57) of ACF67367 from base 3800001 (Photothabdus luminescens nucleoti
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367

Fragment Name	Begin	End
WP ACF67367_00	1	110000
WP ACF67367_01	100001	210000
WP ACF67367_02	200001	310000
WP ACF67367_03	300001	410000
WP ACF67367_04	400001	510000
WP ACF67367_05	500001	610000
WP ACF67367_06	600001	710000
WP ACF67367_07	700001	810000
WP ACF67367_08	800001	910000
WP ACF67367_09	900001	1010000
WP ACF67367_10	1000001	1110000
WP ACF67367_11	1100001	1210000
WP ACF67367_12	1200001	1310000
WP ACF67367_13	1300001	1410000
WP ACF67367_14	1400001	1510000
WP ACF67367_15	1500001	1600000
WP ACF67367_16	1600001	1710000
WP ACF67367_17	1700001	1810000
WP ACF67367_18	1800001	1910000
WP ACF67367_19	1900001	2010000
WP ACF67367_20	2000001	2110000
WP ACF67367_21	2100001	2210000
WP ACF67367_22	2200001	2310000
WP ACF67367_23	2300001	2410000
WP ACF67367_24	2400001	2510000
WP ACF67367_25	2500001	2610000
WP ACF67367_26	2600001	2710000
WP ACF67367_27	2700001	2810000
WP ACF67367_28	2800001	2910000
WP ACF67367_29	2900001	3010000
WP ACF67367_30	3000001	3110000
WP ACF67367_31	3100001	3210000
WP ACF67367_32	3200001	3310000

```
Qy 8 ThrThrArgGlySerProMetProAspThrGlyValLeuArgLeuLeuThrSerGluGln 27
Db 93960 TCTGCGCGAGCCAGCCGCTTCCAGAAATTGGG-----TCGAAAAAATTCGNAAGG 94010
Qy 28 ArgAlaLysGlyTrpArgArgGlnLeuGluGlyLysSerLeuGlyLysPheHisProSer 47
Db 94011 TATTCTGATGTTGG-----GCCTCCGGGTAC-----GGG 94040
Qy 48 -GluThrProTrpIleLysTrpLeuGluGlySerGluThrTrpLysLysValLysLeuPr 67
Db 94041 TAAGACITTTACTGGCAAAAGCAATTGCTGGT-----GAAGCGAAAGTGCC 94085
Qy 67 oThrAspGlyIleSerAlaSer-----LysIleLeuGlyLysIleMetAlaAr 83
Db 94086 GTTCTTTTACTATCTCTGTTCTGATTTTGTGAATATGTTTGTGGTGTGGTGCATCCCG 94145
Qy 83 qValArgIleAlaThrAlaLeuAlaValLeuAlaAlaProCysLeuAlaPhe---As 102
Db 94146 TGTTCGTGATATGTTTCGACAGCGG---AAGAAAGACGACCTTGTATCATCTTTATTGA 94202
Qy 102 pGluValThrAla-----SerGlyValPheProGluGluHi 114
Db 94203 TGAATTTGATGCGGTGGCGCTCAGCGTGGCGCTGTTTGGGTGGCGGTACGATGAACG 94262
Qy 114 sLysHisThr-----GlyGluGlyAr 121
Db 94263 TGAGCAGACTCTAAACCAGATGTTGTTGAGATGGATGGTTTGAAGCAACGAGGTAT 94322
Qy 121 gHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThr-----AlaProAs 139
Db 94323 TATCGTTATTGCGCGGACTAACCGTCCGACGTACTTGAATCCTCGGTGTTGCGTCCAGG 94382
Qy 139 nArgArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArg----- 157
Db 94383 TCGTTTTGACCGTCAGGTGTTGTTGTTGTTGCCAGACGTACGTGGCGCGCAGCATTC 94442
Qy 158 -----AspGlyThrGlyThrValCysGlnIleThrAs 168
```



```

Qy 213 -----ProAlaGlyAspPro----AlaGlyArgLeuVal-----AenProTh 225
Db 30253 TGCATTATCATGAAGCAGCGCATGCTATTATTATGGCGCTGTTCTCTGAGCATGCTGT 30194
Qy 225 rAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVa 245
Db 30193 TCATAAAGTTACCATTTTCGCGTGTGCGGCTTTGGGGTGACTTTTTTCTTACCGGA 30134
Qy 245 lPro-----ThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTr 262
Db 30133 AGGCGATCAGTCAGCTAGTCCGAGAGTTGGAAGCCAGATTTCAACCTTATATGG 30074
Qy 262 pMetAlaLeuAlaArgValProPheMetGlnTyrGlyThrAspGluIleThrTh 282
Db 30073 AGTCGTCTGGCGAAGAGATT-----ATTATGGCCACACAGTGTTTCTACTGG 30023
Qy 282 rAla-----AlaAlaAsnLeuAlaGlyMetGlyG 292
Db 30022 TGCATCTAATGATCAAAAGTGGCGACATCAATTTGCCCGTAACATGGTAACGCAAGTGGG 29963
Qy 292 yPhe-----ProAsnLeuAspAlaValSerIleGlySerAspGlyThrVa 307
Db 29962 TTTCTCGAATAAATTGGCCCTTACTTCTATGCA-----GAGGAAGAAGGGGAAGT 29912
Qy 307 lAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheVa 327
Db 29911 GTTCTTGGCGGTTTCAGTTGCTAAAGCTAAGCACATGCTCATGAATCTCTCTTAT 29852
Qy 327 lSer-----GlnLeuLeuVa 332
Db 29851 TGATCAGGAAGTTAAGCAATTATTCGATCATCAACTATCAACGCTGCTCGTCAGATTCTGAT 29792
Qy 332 lAsnSerPheThrIle-----AspAlaIleThrValGluProLysGlnG 347
Db 29791 GGACAACTTGATGTTCTGCAATTCGATGCAAGATGCAATGATG-----AAGTACGA 29741
Qy 347 uThrPhe---AlaProAspLeuAsnTyrMetValPheAspGluTrpLeuAsnIleG 365
Db 29740 AACTATTGATGCTCCCGAGATT-----GATGACTTAATGATCGTAC 29699
Qy 366 nAsnGlyGlyProProAlaGlyProGluGlu-----LeuAspGluGluLeuArgPh 383
Db 29698 AAATGTTCTGTCACAGCAGGATGGGAAAGTGTATAACGCGACAGTAATAACAGTCGCC 29639
Qy 383 eIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAl 403
Db 29638 AACAGGTAATACACTCTCTCCACAAAGCTGCAAAACCAACTGATGATATCAATATGGC 29579
Qy 403 aTyrArgGlySerLeuIle 409
Db 29578 GCAAGGTGGTGAGCAAGTA 29560

RESULT 15
ID AAF61025 standard; DNA; 3426 BP.
AC AAF61025;
XX
XX
XX
XX 16-MAY-2001 (first entry)
XX
XX P. putida KT2440-associated DNA ORF04882.
XX
XX Transgenic plant; detection; probe; amplification; vaccine carrier;
XX microbial production strain; biological remediation; ds.
XX
XX Pseudomonas putida.
XX
XX DE19935088-A1.
XX
XX 01-FEB-2001.
XX
XX 27-JUL-1999; 99DE-01035088.

```

```

XX
PR 27-JUL-1999; 99DE-01035088.
XX
XX (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKF2-) DKF2 DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX MPI; 2001-192469/20.
XX
XX New DNA sequences specific for Pseudomonas putida KT2440, useful as safe
XX Genetic engineering host, allow detection in presence of other related
XX bacteria.
XX
XX Claim 1a; Page 73-75; 158pp; German.
XX
XX This invention describes novel DNA sequences (I) for specific detection
XX of Pseudomonas putida KT2440. The invention also describes (1)
XX recombinant expression vector containing (1); (2) prokaryotic or
XX eukaryotic cells transfected or transfected with (1) or the vector of (1)
XX; (3) production of expression products by culturing cells of (2); (4)
XX expression products, or their fragments, of (1) and synthetic proteins or
XX peptides with the same sequences (A); (5) poly- or mono-clonal antibodies
XX (Ab) that react specifically with (A); (6) hybridoma cells that produce
XX the monoclonal Ab of (5); (7) transgenic plants that contain transformed
XX or transfected cells of (2); (8) detecting KT2440 using a labeled (1) or
XX Ab as probe; and (9) DNA chips carrying one or more (1). (1), and their
XX fragments, are used as probes to detect and isolate full-length cDNAs
XX and/or to amplify such cDNAs by polymerase chain reaction, and for
XX production of transgenic plants. (1), or antibodies that recognize their
XX expression products, are used for detecting the presence of KT2440,
XX particularly in presence of other, even closely related, bacteria. KT2440
XX is one of the bacteria classified as safe, by the National Institutes of
XX Health, for genetic engineering work, e.g. as microbial production
XX strains, for biological remediation and as vaccine carriers. (1) are
XX exclusive to KT2440 with no significant homology with sequences in other
XX bacteria (specifically the closely related pathogen P. aeruginosa).
XX Compared with other 'safe' bacteria, it has greater catabolic activity
XX and better survival in, and adaptation to, the rhizosphere and soil
XX
XX SQ Sequence 3426 BP; 596 A; 1143 C; 1090 G; 593 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.: 1.31 Length: 3426
Score: 118.00 Matches: 146
Percent Similarity: 31.59% Conservative: 59
Best Local Similarity: 22.50% Mismatches: 230
Query Match: 3.34% Indels: 217
DB: 4 Gaps: 31

US-10-691-383-2 (1-676) x AAF61025 (1-3426)

Qy 111 ProGluHisLysHisThrGlyGluArgHisLeuGln-ThrCysThrAsn----- 128
Db 684 CCTGATCGGAGCGGCACACTTGTGCGGCGGCACTTCAACGCTGATGATGGCC 743
Qy 129 -----SerAspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPh 146
Db 744 GCAGATCAACAGCGAAACCTTGTATGTACCTGACTGACGCGCGCGGTATGTCTCG-- 801
Qy 146 eAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnI 166
Db 802 -----GCGCTGGCGCTGTCTACTGCGACCGACGCACTGCACCACCTG----- 843
Qy 166 eThrAsnGlyGluThrAspLeuAla-----ThrMetPheHisLysSe 180
Db 844 -----GCGCGACCGACCTGCTATCAATGACAGGGTAACACACCTTACCGCTCGTGC 896
Qy 180 rLeuProHisAspGlu-----LeuGlyGlnValTh 190
Db 897 CGAGCCCGGCTCGACGGTTTCCGTGCGAGCGCAGGTGGCGTATTGCTGGGCACTCGGCT 956

```

```
Qy 190 rAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysG1 210
Db 957 GCGCGGT-----GCGAGCGCGCAGTTACGATACCT 989
Qy 210 uAspValProAlaGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaI 230
Db 990 GCAACCGCCACAAGCGAC-----GGCAGGCCCTGGAGATCAGCGCGCAGACGGCG 1043
Qy 230 eAspIleSerGlyProAlaPheSerAlaThrIleProVal-----ProThrLe 248
Db 1044 TGGCAATATCTCGCCAGTTCGACATCAGCGG-----CCGACGTCGACCAACCGATAC 1100
Qy 248 uSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAs 268
Db 1101 CACAGCCCTGAAACAACGACCGACCTG-----GCCCTGGCCGATGG 1142
Qy 268 pValProPheMetGlnTyrGlyThrAspGluIleThr-----ThrAlaAlaI 285
Db 1143 CGTCACCTTACCGGACGCGGTGAGCTGTGTGTACCGTCGAGGTGCGGATGTCGCGG 1202
Qy 285 aAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspG1 305
Db 1203 TAACCTGATTGGCACCGGT-----GTCTCGGTGCGGATGG 1238
Qy 305 yThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyPr 325
Db 1239 C-----CTGTTACGCTCACCTCTGCTGCTGCCCGACCC----- 1272
Qy 325 oPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleThrValGluProLy 345
Db 1273 -----AATGTTGAAGCACTCGACGTGCGTCTGTGTGGATGCGC 1310
Qy 345 sGlnGluThrPheAlaPro-----AspLeuAsnTyrMetValAspPheAspGluTr 362
Db 1311 CGGCAACAGCTCTGCCCCATTGCAATTCGATCGCCCGGACATCACACCGCTGACGCG 1370
Qy 362 pLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuAr 382
Db 1371 AACCAATATTACATTGGTGTGACGG-----CTTGGCCCTGAGCGGTGCGGCGGACCGGTG 1429
Qy 382 gPheIleArgAsnAlaArg-----AspLeuAlaArgValSerPheValAs 397
Db 1430 CCACTGTGGAAGTGGCGGATGCCAATGCACTGTGTGTCG-----GCACGGGTGTCGTGGTGC 1488
Qy 397 pAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSe 417
Db 1489 -AAC-----GGCACCTTCTGATCGACCTCGCCCGCCCGCGC 1523
Qy 417 rArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPh 437
Db 1524 ACAGCCGGCGGAGCAGTTGAGCTGTGTGTCAGACCGAC-----CCAAGCGTAATGCTTCAGT 1580
Qy 437 eGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlu----- 450
Db 1581 GGCTACCGGATGACGATGTACCACTGACCAACCGCGCCGACAGCCCGGCAATCTGGCCAT 1640
Qy 451 -----LeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHi 464
Db 1641 CGATGCTGATGGCACCACTTCCACCGGTACCGCGCTGCGCGCGCTGGCAGCGCTGGARGTGCA 1700
Qy 464 sArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLe 484
Db 1701 T-----GAYGCCAACGGCACGTTGATTGGCAGTGCCATTGCCAACGC 1742
Qy 484 uAspAlaAspPheAspIleSerLeu-----LeuGluAsnAspGluLeuLeuIlyAr 501
Db 1743 CGATGGCAGCTTCAGCATCGAACTCGAACTGAATCCCGCACAAAGCCAAACCGCAACTGCTGGACGT 1802
Qy 501 gValala----- 503
Db 1803 GGTGCCATCATGATACGCGCGCGGTTCCTGCTGCGCGCACAGATCACCGCGCCCGCAT 1862
Qy 504 -----GluIleAsnAl 507
```

```
Db 1863 CACCGCCCTCGCGCACCGACCGAACTGGCGGTGAGCGCGGACCGGACGCTTATCACCGG 1922
Qy 507 aAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySe 527
Db 1923 CCGTGGCGAAACCGGGCAGCACCGTGGCATCGTGGCTGCGAGCGGAACCGAACTGGGCAC 1982
Qy 527 r-----ProThr-----HisProSerTyrProSerG1 536
Db 1983 TGGCGTGGTGGCGCGACCGCGCGTGTTCAGCTCAACCTCAACCCCGCACCAAGTTGACGG 2042
Qy 536 yHis-----AlaThrGlnAsnGlyAlaPheAlaThrValLeuIlyAlaLe 551
Db 2043 CGAAGTGTCTCAAGCCACCGCGACCGATGCGCTGGCAACACTTCCGTTACAGCTCGGT 2102
Qy 551 uIleGlyLeuAsp-----ArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 570
Db 2103 CACCGCGCGGATATCGATGGCGCGATACACCCCG-----CCTGAAGCACCGACCAAT-- 2157
Qy 570 pGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLy 590
Db 2158 -----CTGCTGATCGCCCTTCCGCGCACCGACTCAGTGTGCGAGCGGA 2202
Qy 590 sLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspG1 610
Db 2202 ----- 2202
Qy 610 yIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMe 630
Db 2203 -----GCCGCGCACTACCGTGCAGGTGCGTGACCGCGCGCGCAACATC-- 2244
Qy 630 tThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLe 650
Db 2245 -----CTCGCGACCGGTACCGTCCGCC----- 2265
Qy 650 uPheGlnAspGlyThrPheSerIle 658
Db 2266 ----GCGGATGGCACCTTCACCGTC 2286
```

Search completed: September 19, 2004, 04:26:20
Job time : 1105.36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 03:58:58 ; Search time 8460.83 Seconds
(without alignments)
3463.005 Million cell updates/sec

Title: US-10-691-383-2
Perfect score: 3528
Sequence: 1 MLCRAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVDACQA 676

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10691383/runat_17092004_102656_1741/app query fasta_1.1230
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCI=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10691383 @CGN 1.1 6663 @runat_17092004_102656_1741 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.ey.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

SUMMARIES			
Result No.	Score	Query Match Length	ID
1	3528	100.0	2931 6 AR152190
2	3528	100.0	2931 6 BD235853
3	3528	100.0	2931 6 AR435585
4	3528	100.0	2931 8 AF053411
5	1135.5	32.2	3379 8 LD1491786
6	1129.5	32.0	3420 8 LD1491787
7	626	17.7	1797 6 AX201842
8	626	17.7	1908 8 AF218810
9	626	17.7	2035 8 D87657
10	623	17.7	1794 6 E17199
11	617.5	17.5	2029 8 D87658
12	614.5	17.4	1791 6 E17200
13	135.5	3.8	293350 1 BX294153
14	132	3.7	259950 1 BX294136
15	131	3.7	9579 6 AX598603
16	131	3.7	11018 1 AE011953
17	131	3.7	52101 6 AX598593
18	131	3.7	340750 1 BX294135
19	130.5	3.7	297850 1 AP065577
20	130.5	3.7	302550 1 BX294137
21	130	3.7	1499 8 AK110829
22	129.5	3.7	11679 6 AR318827
23	129.5	3.7	16069 1 AE005097
24	129	3.7	347800 1 AP000060
25	128	3.6	12639 1 AP005072
26	128	3.6	217594 1 AP002086
27	126.5	3.6	13522 1 AE007219
28	125.5	3.6	299650 1 AP005955
29	124.5	3.5	7185 6 AX803764
30	124.5	3.5	11546 1 AE009336
31	124.5	3.5	13971 1 AE008279
32	124.5	3.5	61944 6 AX803750
33	124	3.5	300861 1 AE016777
34	123.5	3.5	10910 6 AX024276
35	123.5	3.5	10945 6 BD262982
36	123.5	3.5	68750 1 AF210843
37	123.5	3.5	68750 6 AR133029
38	123.5	3.5	68750 6 AR199551
39	123.5	3.5	68750 6 AR199559
40	123.5	3.5	68750 6 AR199567
41	123.5	3.5	68750 6 AR201097
42	123.5	3.5	68750 6 AR208671
43	123.5	3.5	68750 6 BD218234
44	123.5	3.5	303450 1 AP005085
45	123	3.5	5820 6 AX616791

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1


```

Db 1968 TGCCTTACATATGAGGGAGAGATCAACAAAGCTCCGGTCAACGCTCGCATTTGGGAGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTTTCGACGGTATCCAGGCCTACTTCTCGGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
Db 2088 ACTGTAGCAACATTCACAGGAGCTGATGACGTTCCGCGAGGAAGCACCTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleIleLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACGGAGAGGTCAATCAAACTTTTCAGACGGGACATTTCTCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTGGTTTACACTGGCGTGGGAGTCCAGGCT 2255

RESULT 2
LOCUS BD235853 2931 bp DNA linear PAT 17-JUL-2003
DEFINITION Gene-modified vanadium haloperoxidase and utilization thereof.
ACCESSION BD235853
VERSION BD235853.1 GI:33045623
KEYWORDS JP 2002525046-A/1.
SOURCE Fucus gardneri
ORGANISM Fucus gardneri
REFERENCE 1 (bases 1 to 2931)
AUTHORS Vreeland,V. and Ng,K.L.
TITLE Gene-modified vanadium haloperoxidase and utilization thereof
JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT OS Fucus gardneri
PN JP 2002525046-A/1
PD 13-AUG-2002
PF 27-AUG-1999 JP 2000570298
PR 10-SEP-1998 US 09/151189
PI VALERIE VREELAND,KWAN L NG
PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/02,C12N15/
PC 00,C12N5/00
CC Gene-modified vanadium haloperoxidase and utilization thereof
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
FEATURES
source 1..2931
/organism="Fucus gardneri"
/mol_type="genomic DNA"
/db_xref="taxon:3013"

ORIGIN
Alignment Scores:
Pred. No.: 2,49e-249 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2 (1-676) x BD235853 (1-2931)

Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
Db 228 ATGCTTTGCCATGCGCGGACACGACGAGGCTCTCTATGCTCTGACCGGAGTGTCTT 287
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaGlyGlyTrpArgArgGlnLeuGluGlyGluLys 40
Db 288 CGGTGTCTCATCAGACGACGGCGCTAAAGTTGGAGACGCCAGATTAGAGGGGAGAAA 347
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleIleYsrLeuGluGlySerGluThr 60
Db 348 TCACTAGGTTTTCATCCAAGCGGACGCGCTTATATCAAGTACTTTGGAAGGCTCTGAGACT 407

```

```

Qy 61 TrpLysLysValIysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
Db 408 TGAAGAAGGTAAAGCTTCCAAACGACGGCATATCGGCTTCCAAAGATCTCTGGGTAAATTT 467
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db 468 ATGCCCAAGGTCGCGCATCGCTACCGCTTGGCTGTGGTACTGGCCGACACCTCTGTTGGCA 527
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
Db 528 TTGACGAGGTCACAGCGCAGTGGTGTGTTTCCCTGAGGACACCAAGACACACCGGGAGGGA 587
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArg 140
Db 588 AGACACCTCCAGACCTGTGTAACAACTCCGACGATCGCTGGATCCGACGGCCGCAATAGA 647
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 648 AGGACACACCTAGCTTTTGGCTCGCGCGCGCATGCCCGCAGGCGAGAACGTCACGGGACA 707
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
Db 708 GGGACTGTCTGCCAAATCACTAACGGAGAAACTGATTTGGCTACCATGTTCCACAGTCT 767
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
Db 768 CTGCCACACGATGAACCTGGGACAGGTAACCCGACAGACACTTCGCTATCTCTCGAGGACTGC 827
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
Db 828 ATCTTAAACGGAGATTTTTCAGCATTTTGGCAGACCGTGGCTGGCGGAGACCCCGCGGTGC 887
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 888 CTGCTCAATCTTACCGCTGGTTGGCATCCGATATCCGCTCCGCAATCTTCGGCTACG 947
Qy 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 948 ACAATACCCCGGTACCTACCTTTCTCTCTGAGCTCGCGCTCAGTTGGCGGAGCTA 1007
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 1008 TACTGGATGGCGTGGCGAGGATGTACCTTTATGCAGTATGGCACCAGCAAAATTACC 1067
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaValSer 300
Db 1068 ACTACCGCGGACCAACCTCGCTGGAAATGGAGGCTTCCAAATCTGGACCGCGTGTCTG 1127
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
Db 1128 ATAGGTTCCGATGATGATCGGTGGACCCGTTCTCCAGCTCTTCCGAGCGACCTTCTGTTGGT 1187
Qy 321 ValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAlaIle 340
Db 1188 GTTGAACCGGGCGCTTTGTCTCTCAGCTGTCTGTGTAACAGCTTTCACCATCGACGCTATT 1247
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAsp 360
Db 1248 ACGTTCGAACCGAAGCAGGAGACATTCGCCCGGACCTTGAACCTATATGTTGCTGATTTGAC 1307
Qy 361 GluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
Db 1308 GAATGGCTGAACATTCAGATGTTGGACCCCGCGCGCGCCCGAAGAGTTAGACGAAGAG 1367
Qy 381 LeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsn 400
Db 1368 CTGCGTTTATTCGGTAAACGCCCGCGACCTGGCCAGGGTCTCTCTTCGTGGCAATATCAAC 1427
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
Db 1428 ACCGAAGCTTATCGCGGTCTCTTTATCTTACTTGGCTGGAGGCTTTCAGCAGGCCCGGT 1487

```



```
Qy 421 ILeAsnGlyProPheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSer 440
Db 1488 ATCAACGGTCAATTCATCGACGATTCGGCAGCGGGCTTCGTCACACTTCGGGACAGTCT 1547
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
Db 1548 CACTACTTCAGATTGATAGTGGTCCCGCAGCTGCGCAGCGTGCCTCGTGTACCAAAAG 1607
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIle 480
Db 1608 TGGCAGGTGTCATCGATTTCACGCGCCGAGGCTCTCGGGGGTACCTCCCAACACCATC 1667
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluLeuAspGluLeuLys 500
Db 1668 GCGGGGATCTAGATGCGAGCTTCGACATCTCCCTTCTTGAAATGATGAGCTCTTGAAA 1727
Qy 501 ArgValAlaGluIleAsnAlaAlaGlnAsnProAsnGlnValAlaThrTyrLeuLeuPro 520
Db 1728 CGTGTGGCGAGATAAATGCGCGCAGAAATCCCAACACGAGGTCACTACCTTCTTCCA 1787
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 1788 CAAGCTATCCAAAGTGGGATCGCCAAACGACCCCTTCTTACCGTCCGCGCACGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
Db 1848 AATGGACATTTGCCACAGTTCTGAGGCCCTCATTCGGCTGAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAla 580
Db 1908 TTCCCTTAACCCCGTGTCCCAAGCGATGACGGCGTGGAACTAATCAACTTCGAAGGGCA 1967
Qy 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
Db 1968 TGCCTTACATATGAGGAGAGATCAACAAGCTCCGCTCAACGTCGATTTGGGAGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTTCGACGGTATCCAGGCCTACTTCTCGAGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
Db 2088 ACTGTACCAACACTTCACAGGAGCTGATGAGTTCGCGGAGGAAGCCACTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTGATCAAACTTTTCAGGACGGGACATTCCTCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGGTTTGGTTTACACTGGCGTGGCGAGTGCAGGCT 2255

RESULT 3
AR435585 2931 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 1 from patent US 6656715.
DEFINITION AR435585
ACCESSION AR435585
VERSION AR435585.1 GI:40198553
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2931)
AUTHORS Vreeland,V.
TITLE Recombinant minimal catalytic vanadium haloperoxidases and their
uses
JOURNAL Patent: US 6656715-A 1 02-DEC-2003;
FEATURES
Location/Qualifiers
1..2931
/mol_type="genomic DNA"
source
ORIGIN
```

```
Alignment Scores:
Pred. No.: 2,49e-249 Length: 2931
Score: 3528.00 Matches: 676
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2 (1-676) x AR435585 (1-2931)
Qy 1 MetLeuCysHisAlaAlaAspThrThrArgGlySerProMetProAspThrGlyValLeu 20
Db 228 ATGCTTTGCCATGCAGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTGCCT 287
Qy 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTrpArgArgGlnLeuGlyGluLys 40
Db 288 CGGTGTGCTCATCAGAGCAGCGCTAAAGGTTGGAGACGCCAGATTAGAGGGGGAGAAA 347
Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60
Db 348 TCACTAGTGTTCATCCAGCGAGACGCCCTTATATCAAGTACTTGGAGAGGCTCTGAGACT 407
Qy 61 TrpLysLysValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80
Db 408 TGAAGAAGGTTAAGCTTCCACCGACGCATATCGGCTTCCAAGATCTCTGGGTAAATTT 467
Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100
Db 468 ATGCCACAGGTCGCGCATCGCTACCGCTTGGCTGTGGTACTGGCGGCACCCCTGTTTGGCA 527
Qy 101 PheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGly 120
Db 528 TTGCAGAGGTTCACAGCCAGTGTGTTCCTTCGAGGAAACACAAGACACACCGGGGAGGA 587
Qy 121 ArgHisLeuGlnThrCysThrAsnSerAspAspAlaLeuAspProThrAlaProAsnArg 140
Db 588 AGACACCTCCAGACCTGTACAAACTCCGACGATCGCTGGATCCGACGGCGCGCAATAGA 647
Qy 141 ArgAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160
Db 648 AGGACAACTAGCTTTTGGTTCGCGCGCGATCGCCCGAGGAGGAGAACGTGACGGGACA 707
Qy 161 GlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSer 180
Db 708 GGACTGCTGCCAATCACTACCGAGAACTGATTTGGCTACCATGTTCCACAAGTCT 767
Qy 181 LeuProHisAspGluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCys 200
Db 768 CTGCCACACGATGAACCTGGGACAGGTAAACCGCAGACGACTTCGCTATCTCTCGAGGACTGC 827
Qy 201 IleLeuAsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg 220
Db 828 ATCTTAAACCGAGATTTCAGCATTTGCGAGGAGTGCCTCGGGAGAGACCCGCGGGTCCG 887
Qy 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
Db 888 CTGTCATATCTACCGCTGCTGTTGCCATCGACATATCCGGTCCCGCAATCTCGGCTACG 947
Qy 241 ThrIleProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
Db 948 ACAATACCCCGGTACCTACCTTCTCTCTGAGCTCGCGCTCGCTTTCAGTTGGCGGAGCTA 1007
Qy 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
Db 1008 TACTGGATGGCGTGGCCAGGATGATACCTTTATGAGTATGTCACCGACCGAATATACC 1067
Qy 281 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSer 300
Db 1068 ACTACCGCGCAGCCAACTTCGCTGGAATGGAGGCTTCCCAATCTGGAGCGCGTCTCG 1127
Qy 301 IleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGly 320
Db 1128 ATAGGGTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187
```

```
Qy 321 ValGluThrGlyProPheValSerGlnLeuLeuValAenSerPheThrIleAspAlaIle 340
Db 1188 GTTGAACCGGCGCCTTTGTCTCTCAGCTGCTCGTGAACAGCTTCCACCATCGACGCTATT 1247
Qy 341 ThrValGluProLysGlnGluThrPheAlaProAspLeuAenTyrMetValAspPheAsp 360
Db 1248 ACGGTCGAACCGAAGCAGGAGACATTCGCCCGCCACTTGAACTATATATGCTCGATTTTGAC 1307
Qy 361 GluTrpLeuAenIleGlnAenGlyGlyProProAlaGlyProGluGluLeuAspGluGlu 380
Db 1308 GAATGGCTGAACATTCAGATGTTGGACCCCGCGCGCCCGCCGAGATTAGACGAAGAG 1367
Qy 381 LeuArgPheIleArgAenAlaArgAspLeuAlaArgValSerPheValAspAenIleAen 400
Db 1368 CTGGCTTTATCCGTAACCGCGCCGACCTCGCCAGGGCTCTCTTCGTGGCAATATCAAC 1427
Qy 401 ThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGly 420
Db 1428 ACCGAAGCTTATCGCGGCTCTCTTATCTTCTTATCTTGTAGCTGGAGGCTTTCAGACGCCCGGT 1487
Qy 421 IleAenGlyProPheIleAspSerArgGlnAlaGlyPheValAenPheGlyThrSer 440
Db 1488 ATCAACGGTTCATTCGACAGTATCGGACGGCGGCTTCGTCAACTTCGGCACGCTCT 1547
Qy 441 HisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLys 460
Db 1548 CACTACTTCAGATTGATAGTGCCTGCGCGAGCTGCGCGAGCTGCTCGTGTACCAAAAG 1607
Qy 461 TrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAenThrIle 480
Db 1608 TGGCAGGTGATCGATTGTCAGCGCCCGCGAGCTCTCGGGGGTACCTTCCACACACATC 1667
Qy 481 AlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAenAspGluLeuLys 500
Db 1668 GCGGGGATCTAGATGCGACTTCGACATCTCCCTCTTGGAAATGATAGCTCTTGAAA 1727
Qy 501 ArgValAlaGluIleAenAlaAlaGlnAenProAenAenGluValThrTyrLeuLeuPro 520
Db 1728 CGTGTGGCGAGATAAATCGCGCGAGATCCCAACAGAGGTACCTACCTTCTTCCA 1787
Qy 521 GlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGln 540
Db 1788 CAAGCTATCCAAAGTGGATCGCAACAGCACCCTTCTTACCCGTCGCGCCACGCTACCCAA 1847
Qy 541 AsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCys 560
Db 1848 AATGGACATTTGCCACAGTTCTGAAGCCCTCATTTGCCCTAGATCGGGAGGTGAGTGC 1907
Qy 561 PheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAenPheGluGlyAla 580
Db 1908 TTCCCTAACCCCGTGTTCCTCAAGCGATGACGGCTCGAATCACTTTCGAAGGGGCA 1967
Qy 581 CysLeuThrTyrGluGlyGluIleAenLysLeuAlaValAenValAlaPheGlyArgGln 600
Db 1968 TGCCTTACATATGAGGAGAGATCAACAAAGCTCGCGTCAACGTCGATTTGGAGGCAG 2027
Qy 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIle 620
Db 2028 ATGCTGGGCATCCACTATCGGTCGAGTATCCAGGCTTACTTCTCGAGAGACAAATC 2087
Qy 621 ThrValArgThrLeuHisGlnLeuMetThrPheAlaGluGluAlaThrPheGluPhe 640
Db 2088 ACTGTACGAACATTCACAGGAGCTGATGACGTTCCCGCGGAGAACCCCTTTGAATTC 2147
Qy 641 ArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGly 660
Db 2148 CGCTTATTCACCGAGAGGTGATCAAACTTTCCAGGACGGGACATTTCTCCATCGATGGA 2207
Qy 661 AspMetCysSerGlyLeuValTyrThrGlyValAlaAspCysGlnAla 676
Db 2208 GATATGTGTTCCGCTTTTGGTTTACACTGGCGGTGGCGAGCTGCGCAGGCT 2255
```

RESULT 4

AF053411

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

5'UTR

CUS

2931 bp mRNA linear PLN 12-SEP-1998
Fucus distichus vanadium bromoperoxidase mRNA, complete cds.

AF053411.1 GI:3582762

Fucus distichus

Fucus distichus

Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.

1 (bases 1 to 2931)

Vreeland, V.; Ng, K. and Epstein, L.

Direct Submission

Submitted (11-MAR-1998) Environmental Science, Policy and

Management, University of California, 201 Wellman Hall, Berkeley,

CA 94720-3112, USA

Location/Qualifiers

1..2931

/organism="Fucus distichus"

/mol_type="mRNA"

/db_xref="taxon:3012"

/dev_stage="2-cell embryo"

1..227

228..2258

/codon_start=1

/product="vanadium bromoperoxidase"

/protein_id="AAC35279.1"

/db_xref="GI:3582763"

/translational="MLCHAADTTGSPMPDTGVLRLTSEQRAGKWRQLEGEKSLGP

HPESTYIKYLEGSEWKVKLPTDGISASKILGKIMARVRIATALAVLAAPCLAFD

EVTSAGVPEEHKHTGEGRHQCTNSDDALDPTANRNNDVAFASRRDAARERDGT

GTVCCTNGETDLATMFHKSPLDHEQVTDADDFAILEDCLNGDFSCIDEPAGDPA

GRLVNPTAAFAIDISGPAFASATVPVPTLSSPELAQLAELYMALARDVPFMQYGT

DEITTTAAANLAGMGFPNLDVSIISGSDGTVDFEOLFRATFVGVETGPPVSOLLVNS

FTDALTVPKQSTFAPDLNMYVDFDEMLNIONGPPAGPELDEELRFRINARDLAR

VSFDNINTEAYNGSLILLELGAFSRPGINGPIDSQDQAGFVNFSTSHYFRILIGAE

LAQASCYQKVRHFRARPEALGTLHNTIAGDLADDFISLLEDELLKRVAEINAA

QPNNVTVLLPQAIQVSPSTHPSGHATQNGAFATVLKALIGLDGCGECPNPVF

PSDGLIELFEGACLTVEGINKLVANVAFGRQMLGIHYRFDGIQGLLIGETITVRT

LHQELMTFAEATFEFLFTGEVILKFDQGTFSIDGDMCSGLVTVTVADCOA"

2259..2931

3'UTR

ORIGIN

Alignment Scores:

Pred. No.:	2,498-249	Length:	2931
Score:	3528.00	Matches:	676
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-691-383-2 (1-676) x AF053411 (1-2931)

Qy 1 MetLeuGlyHisAlaAlaAspThrArgGlySerProMetProAspThrGlyValLeu 20

Db 228 ATGCTTTGCCATCGCGGACACGACAAAGAGGCTCTCTATGCTGACACCGGAGTCTT 287

Qy 21 ArgLeuLeuThrSerGluGlnArgAlaLysGlyTyrArgArgGlnLeuGluGlyLys 40

Db 288 CGGTTGCTTCATCAGACGACGCGCTAAAGTTGGAGACGCCAGCTTAGAGGGGAGAAA 347

Qy 41 SerLeuGlyPheHisProSerGluThrProTyrIleLysTyrLeuGluGlySerGluThr 60

Db 348 TCACTAGTTCATCCAGCGAGACGCTTATATCAAGTACTTGAAGGCTCTGAGACT 407

Qy 61 TrpLysIleValLysLeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIle 80

Db 408 TGAAGAAGGTTAAGCTTCCACGCGGACGATATCGGCTTCCAGATCTCCGGGTAAATTT 467

Qy 81 MetAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAla 100

Db 468 ATGCGCCAGGGTCCGCATCGCTACCGCTTGGCTGTGTGTACTGGCCGCCACCTGTTGGCA 527

QY 582 -----LeuThrTyrGluGlyGluIleAsnLeu 591
 Db 1608 GTCAACGAGAGGCGAGCTCGTGAAGGCTCAGTACGAGGAGAGCTCAACAGATC 1667
 QY 592 AlavalAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle 611
 Db 1668 AGCCCAACCTGCTCTTGGGAAGTCCACATCGTGTTCACCTGGAGGATGGAGCGGTG 1727
 QY 612 GlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631
 Db 1728 TACGAGAGCGCTCATGGAGAGACAGCTGCGTCCGCGCTCGAGCAGGAACCTTCCTGGC 1787
 QY 632 PheAlaGlu-----GluAlaThr 637
 Db 1788 CTCCCGAGGCGCGAGGTGAGGGCAAGAGCGCGTGTGACATCCCGCGGCGACT 1847
 QY 638 PheGluPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSer 657
 Db 1848 TACAATTCGCGCTTTACAGCGGTAAAGATTCTTGAGCTCTACGCGAGGAACCTTGTAACAG 1907
 QY 658 IleAspGlyAspMetCysSerGlyLeuValTyrThrGlyValAlaAspCys 674
 Db 1908 CTGGACGCAAGCTCTCGAGGGG---GCCTTCACGCGGACGACTTCTGT 1955

RESULT 6

LOCUS LDI491787 3420 bp mRNA linear PLN 28-JUN-2003
 DEFINITION Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 2 (VBPO2 gene).

ACCESSION AJ491787.1 GI:32329413

VERSION vanadium-dependent bromoperoxidase 2; VBPO2 gene.

KEYWORDS Laminaria digitata

SOURCE Laminaria

ORGANISM Eukaryota; stramenopiles; Phaeophyceae; Laminariales;

Laminariaceae; Laminaria.

REFERENCE

AUTHORS Colin,C., Leblanc,C., Wagner,E., Delage,L., Leize-Wagner,E., Van Dorselaer,A., Kloareg,B. and Potin,P.
 TITLE The brown algal kelp Laminaria digitata features distinct bromoperoxidase and iodoperoxidase activities
 JOURNAL J. Biol. Chem. 278 (26), 23545-23552 (2003)
 MEDLINE 22703831
 PUBMED 12697758
 REFERENCE 2 (bases 1 to 3420)
 AUTHORS Leblanc,C.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUN-2002) Leblanc C., UMR 1931, CNRS-Goemar, Station Biologique, BP 74, 29682 Roscoff, FRANCE

FEATURES

source
 1..3420
 /organism="Laminaria digitata"
 /mol_type="mRNA"
 /db_xref="taxon:80365"
 /tissue_type="sporophyte"
 /clone_lib="lambda ZAP II - S7"
 /country="France: Brittany"
 1..3420
 /gene="VBPO2"
 1..36
 /gene="VBPO2"
 37..2085
 /gene="VBPO2"
 /codon_start=1
 /product="vanadium-dependent bromoperoxidase 2"
 /protein_id="CAD37192.1"
 /db_xref="GI:32329414"
 /translation="MKPARPTQCALYSVLRLHVALCVACAFLLFEYQISSLGL
 TALLAPAPCGHEEPEPTQPLLSGNVCRVRSQDLDPVPRKVTLLKRLALAIKDE
 ISVGTCHVNGDEENVLPFAGQYKTLPHDKFGQVDEDAKLLKLCVRSIDNCEK
 VPSGARGGAKLTNPFGTAHQVTGADSDNVITTPDSLLSLRAAQAEVYVMALL
 RDPFGFEAKNDYVRLAAENLQSLPAPKGLNIPRSEGGKIDPVDLPRITWPVGVTTPG

gene

5'UTR

CDS

3'UTR

2086..3420

/gene="VBPO2"

ORIGIN

Alignment Scores:

Pred. No.: 3,41e-73 Length: 3420
 Score: 1129.50 Matches: 253
 Percent Similarity: 51.82% Conservative: 102
 Best Local Similarity: 36.93% Mismatches: 247
 Query Match: 32.02% Indels: 83
 DB: 8 Gaps: 14

US-10-691-383-2 (1-676) x LDI491787 (1-3420)

QY 51 TyrIleLysTyrLeuGluGlySerGluThrTyrPheLysValLysLeuProThrAspGly 70
 Db 7 CATACCAAGAGTTTAGACACACAGCGCTCGCCATGAAGCGCTAGCCACCCAGCCCGGC 66
 QY 71 IleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArgIleAlaThrAlaLeu 90
 Db 67 -----GGCGTGTACTGTCTCTCTCTCTCCGACACGTTGGGTGTGCCCTG 111
 QY 91 AlaValValLeuAla----- 95
 Db 112 TGGTGTGTCATCGCGCTTTCCTTTGTTGCAATATCAGATTTCCTCTTGGGCTTGACC 171
 QY 96 -----AlaProCysLeuAlaPheAspGluValThrAlaSerGlyVal 109
 Db 172 GCTTTGTGCGCGGCTTTCGCTCGCTCGCTCGAGGAG----- 213
 QY 110 PheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSer 129
 Db 214 ---CCTCCGAGCCCGAGCGCTCTGCTCAGCGGGAACGTGTGCAGGTGCGAGACTCC 270
 QY 130 AspAspAlaLeuAspProThrAlaProAsnArgArgAspAsnValAlaPheAlaSerArg 149
 Db 271 CTCGACTTCTTCACCGCGGTGCGC-----CGCGGAAGGTGACCTCTCTC---AAG 318
 QY 150 ArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIleThrAsnGly 169
 Db 319 CGCTGCGCCATCGCAAGGACGAGATATCGGTTGGACCCCACTTGCACAGCAACAGCGC 378
 QY 170 GluThrAsp-----LeuAlaThrMetPheHisLysSerLeuProHisAspGlu 185
 Db 379 GACAGGAAACGTCGCGCTCTTCGCTGGCGAGTACCAAGAGCGCTACCGCAGCAAG 438
 QY 186 LeuGlyGlnValThrAlaAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAsp 205
 Db 439 TTCGTCAGGTGGACGAGCATACAGAAGCTGCTGGAGTGCCTCTTCCAGCAGCGAC 498
 QY 206 PheSerIleCysGluAspValProAlaGlyAspProAlaGlyArg----- 220
 Db 499 ATCAACGAGTGCAGAGGTGCGCTCTCGGC-----GCCGCGCAAGAGGTGGCGCGAAG 552
 QY 221 LeuValAsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr 240
 Db 553 CTGACCAACCCCTTCGGGGGACCGCGCACAGGTACCGGTGCTGACAGCGCAATGTG 612
 QY 241 ThrIleProProValProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeu 260
 Db 613 TTCATCACCCCGCCGACAGTCTCTCTCCGAAAGGTGCGCGCGCAGCAGCAAGATT 672
 QY 261 TyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIleThr 280
 Db 673 TACTGGATGGCCCTGTGTTAGGACATCCCTTCCGGGAGTTTCGCGAAGAACTACTCGTC 732

VVSQFMLSDFLIDSIKVTAKADPLTPEVDYMTAFQFWLDVQNGASKLETTDFBENPRF
 IRNRDLATIALRDOLYTFEAFRAALILFTFEGALGGEVGPYAEAEKQCFATFGEPHIL
 TAMASSTRHAWYAKVQVHRMLRPEALVHNTLMRDVITPLPDSILRNTELLNR
 VEHNRMPDGEKTELLPMAAAGSPTHPAYPSHAINNGAVITALKALFLYEAQOK
 CFPNVSNDEGTIKRIKPSGREIVGECVNEKGLVEGTIGELNKNISANVILGRS
 HIGVHRMVGVALMGETSCVRLQLOELPGLPEAREVEGKRRGDIPPATYKFLYLS
 GKILELYGNLYKLDGLKCEGAFGTGDFDIDEDDYSSFFDIIVEEHAQFSLHGTSL
 "

QY 266 AlaArgAspValProPhe-----MetGlnTyr 274
|||:||||:||||
Db 508 GGAGCGCAATCGAGTTTACGAGTTGATCCCAAGAACGACGAGATATATCAGTT 567
QY 275 GlyThrAspGluIleThrThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
|||:||||:||||
Db 568 GCTATTGATCAGCTTAACGGGTAGAGTGGTTCAACACGCGCGCAATGCTCGGAGATCCG 627
QY 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
|||:||||:||||
Db 628 CCTCGCAAAATCGCTCGCTCGCGGTGAGTGAAGTGC-----GGAACATTGTTTC 678
QY 315 ArgAlaThrPheValGlyValGlnThrGlyProPheValSerGlnLeuLeuVal----- 332
|||:||||:||||
Db 679 CGCGGTATTCTTCAGGCTCTGAGGTGCGCCCGTATCTCAGCCAGTACATCATCGTTGGT 738
QY 332 ----- 332
Db 739 AGCAGCAGATTGGCTCCGCGACGGGTGGCAACAAACTCTCGTACGCCCCAATGCCGCT 798
QY 333 -----AenSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
|||:||||:||||
Db 799 GATGAGTTTGTATGCGTAAATCGCTACGGAAGCATCACCATTAGCCAGCGGTGCGTATC 858
QY 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
|||:||||:||||
Db 859 GCCACGCTCGGGCGGACTTTATGACCGATTGAGGTATCTTGACGTCGACGAGCT 918
QY 369 GlyProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
|||:||||:||||
Db 919 CGGAGCTTCGAGGCTTTGATGCTATGAGCGCGGACGCGCTCATCCGACGATCCGC 978
QY 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
|||:||||:||||
Db 979 GATCTTCGACGCTGGGTGACCTTTGACGCACTGTACGAGGCTACCTCAATGCGTGCCTG 1038
QY 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
|||:||||:||||
Db 1039 ATTCTGTGGCGAACCGCGTGGTTCGATCCCAACATT-----CGTTCCAGCAGGAG 1092
QY 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
|||:||||:||||
Db 1093 GACAAGCTCGATAACACGAGCGCTTTGTGAACCTTCGAGACGACACGCTGAGTCTG 1152
QY 446 Ile-----GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHis 464
|||:||||:||||
Db 1153 GTGACTGAGGTGGCTACGCGCGCTTGAGCGGTACGCTACGAGAGTTTAACTTCAT 1212
QY 465 ArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHis-----AsnThrIleAlaGlyAsp 483
|||:||||:||||
Db 1213 CGTCCGCTCGGCTGAGGCTACCGGTGGTCTGATTAGCTGACAAACAAATCCGACGGAG 1272
QY 484 -----LeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 499
|||:||||:||||
Db 1273 AAGGGCAGAGCGTTTCCCTGAGTTGACTTCTGTTGAGAGCTTGAAGATATCTG 1332
QY 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
|||:||||:||||
Db 1333 GAGAAAGCTGAAATTAGCAATAGGAAACAGAAATCTGCTGACGAGAGATCTGACCCCTGAT 1392
QY 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534
|||:||||:||||
Db 1393 CCTTCATCTCTGTTCGCGCAGGCATTCGCGAGGCGAGCCCATCTCCATCTGCTACCGA 1452
QY 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly--- 553
|||:||||:||||
Db 1453 AGCGGCCACCGTGTGGTGTGCGCATGTGTGACGATCTCAAGGGTCTTCGACTCC 1512
QY 554 -----LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAsp 570
|||:||||:||||
Db 1513 AACTTCAGATCGATCAGGTGTTTCGAG-----GTGCAAAAGATGAGGAC 1557

571 GlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLys 590
|||:||||:||||
Db 1558 AACCTTGTAAAGTCGCTTTTCAAGGGAAT---CTCACTGTTCCGCTGAGCTGAACAAG 1614
QY 591 LeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGly 610
|||:||||:||||
Db 1615 CTGCGCGCAATATTGCGATCGCGCGGAACATGCGAGGTGTTCTACTACTTCTCTGACCAG 1674
QY 611 IleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630
|||:||||:||||
Db 1675 TTCCGATCAATTCGTTGTTGAGCAGGTTGCGAATCTTGGAAAGACGAGTCTG 1734
QY 631 ThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
|||:||||:||||
Db 1735 ACATATGCGGAGAACTTCTTCAACTTCCGCAAGTTTGTGGAACATCAATCCAGATC 1794
RESULT 8
AF218810 1908 bp mRNA linear PLN 14-AUG-2002
LOCUS Corallina officinalis vanadium-dependent bromoperoxidase mRNA,
DEFINITION complete cds.
ACCESSION AF218810 GI:21309839
VERSION AF218810.1
KEYWORDS
SOURCE Corallina officinalis
ORGANISM Corallina officinalis
Eukaryota; Rhodophyta; Florideophyceae; Corallinales;
Corallinaceae; Corallinoideae; Corallina.
REFERENCE 1 (bases 1 to 1908)
AUTHORS Carter, J.N., Beatty, K.E., Simpson, M.T. and Butler, A.
TITLE Reactivity of recombinant and mutant vanadium bromoperoxidase from
the red alga Corallina officinalis
J. Inorg. Biochem. 91 (1), 59-69 (2002)
JOURNAL 22117781
MEDLINE 12121762
PUBMED
REFERENCE 2 (bases 1 to 1908)
AUTHORS Carter, J.N. and Butler, A.
TITLE Direct Submission
SUBMITTED (23-DEC-1999) Chemistry, University California Santa
Barbara, Santa Barbara, CA 93106, USA
FEATURES
Location/Qualifiers
1..1908
/organism="Corallina officinalis"
/mol_type="mRNA"
/db_xref="taxon:35170"
1..1197
/codon_start=1
/product="vanadium-dependent bromoperoxidase"
/protein_id="AA046061.1"
/db_xref="GI:21309840"
/translation="MGI PADNLOSRAKASFDRVSAELALARGVWVPSLANGBELLYR
NPDENGDSFVSFTKGLPHDDNGAIIDPDELAFVRAINSGDEKEIADLTIGPAD
PDGLPIWSDLEANSLEVRGWENSSAGLTFDLEGPDAQSIAMPAPVLTSPELIAE
IABLYLMALGREIEFSPDKNAEYIQFADQNGLEWNTFAMIGDPPAEIRRRRG
EVTGNLFRGLPGSEVGYLYQIIVGSKQIGSATGKNTLVSPNAADFGEIAYG
STISORVRIATPGDRDFTDKVFLVDQDAADFRGESYEPGARLRTIRDLATWVHF
DALYEAVALNACILLANRVFPDNPFOQEDKLDNDQVFNFGDAHVLSTVETARA
LKAVRIQKNHNRHRLREPATCGLLISVNTKNAEKGESVFPFVDLAVELEILEKAEIS
NRKQIADGDDPDPSPFLUPQAFAGSPHPSYSGSHAVVAGACVILKKAFFDSNFQI
DQVFEDKEDKLKSSFKGTLTAVAGELNKLDNIAIGRNAGVHVPFSDQFESILLGE
QVAIGILEBQSILTYGENFFNLPKFDGTTIQI"

Alignment Scores:
Pred. No.: 1,39e-36 Length: 1908
Score: 626.00 Matches: 176
Percent Similarity: 45.00% Conservative: 76
Best Local Similarity: 31.43% Mismatches: 220
Query Match: 17.74% Indels: 88
DB: 8 Gaps: 15
US-10-691-383-2 (1-676) x AF218810 (1-1908)

QY 168 AsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGly 187
|||||
DB 148 AACGGAGATCAAGCTTCATGCTTACGATGCTTACAAAGGGTCTTCGCGATGACGAAATGGC 207
QY 188 Gln---ValThrAlaAspAspPheAlaIleuGluAspCysIleLeuAsnGlyAspPhe 206
|||
DB 208 GCTATTATGACCGGACGACTTCTGGCCCTTGGTCTGGGATCAATAGTGGAGATGAG 267
QY 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
|||||
DB 268 AAGGAGATCGCGGACCTCACATTGGGGCCAGCTCGCGACCGGACACTGGCTTACCAATC 327
QY 222 -----ValAsnProThr 225
DB 328 TGGCGCTCGGATTGGCGAATTTCTCGAGCTCGAAGTGGGAGGATGGGAAACAGCTCT 387
QY 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245
|||
DB 388 GCGGCTCTACCTTCGACCTTGGAGGCGCGAGCGGAGTGGATGCCATGCCACCGGCG 447
QY 246 ProThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeu 265
|||
DB 448 CTTGTGCTCAGGACCTGAGCTCATCGCGGAGATACGAGACTGTACTGATGGCGCTT 507
QY 266 AlaArgAspValProPhe-----MetGlnTyr 274
|||||
DB 508 GAGCGCGAATTCGAGTTTAGCGAGTTTTCGATTCGCCAAAGAACGACGAGATATATTCAGTTT 567
QY 275 GlyThrAspGluIleThrThrAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
|||
DB 568 GCTATTGATCAGCTTAAACGGGCTAGAGTGGTTCAACACGCGGCAATGCTCGGAGATCCG 627
QY 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
|||
DB 628 CTTGCGGAATCCGTCGCCCTCGCGTGGAGTGTCTC-----GGAACTTGTTC 678
QY 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuVal----- 332
DB 679 CGCGGTATTCTTCAGGCTCTGAGGCTCGGCGCGGTATCTCAGCCAGTACATCATCTGTTGGT 738
QY 332 ----- 332
DB 739 AGCAAGCAGATTGGCTCCGCGACGCGGTGGCAACAAACTCTGTGAGCCCCCAATGCGGCT 798
QY 333 -----AsnSerPheThrIleAspAlaIleThrValGluProLysGlnGluThr 348
|||
DB 799 GATGAGTTTATGGTGAATCGCTACGGAAGCATCACCATTAGCCAGCGGTGCGTATC 858
QY 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
|||
DB 859 GCCACGCTGGGCGGACCTTATGACCGCATTTGAAGGTATCTCTGACGTCGAGGACGCT 918
QY 369 GlyProAlaGlyProGluGluLeuAspGluLeuArgPheIleArgAsnAlaArg 388
|||
DB 919 GCGGACTTCGAGCTTTGAGTCGTATGACCGGAGCAGCGCTCATCCGAGCATCCGC 978
QY 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeu 408
|||
DB 979 GATCTTGGGACGCTGGTGCATTTGACGCACTGTACAGCGCTTACCTCAATGGTGGCTG 1038
QY 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
DB 1039 ATTCTGTGGGAACCGGCTGCGCTTCGATCCCAACATT-----CGTTTCCAGCAGGAG 1092
QY 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
|||
DB 1093 GACAAGCTCGAATACCGAGGACGCTGTTGTGAACCTTCGAGACCGACACGCTGCTGAGTCTG 1152
QY 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHis 464
|||
DB 1153 GTGACTGAGTGGCTAGCGCGGCTGAGCGGCTGACGATCCGATCCAGAGTTTAACTTCAT 1212
QY 465 ArgPheAlaArgProGluAlaLeuGlyThrLeuHis---AsnThrIleAlaGlyAsp 483

DB 1213 CGTGGCTCGCGCTGAGGCTACCGGTGGTCTGATTAGCTGGAACAAATCGCAGCGAG 1272
QY 484 -----LeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 499
DB 1273 AAGGGCGAGAGCGTTTCCCTCGAGTTGACTGCTGCTGTGAGAGAGCTTGAAGATATCTGT 1332
QY 500 LysArgValAlaGluLeuAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
DB 1333 GAGAAAGCTGAAATTAGCAATAGGAAACAGAACATTGCTGACGGAGATCTCTGACCTGAT 1392
QY 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534
DB 1393 CTTTCATTTCTGTTCGCGCAGGCATTCGCCGAGGCGAGCCCATTCCTCCGCTCTACGGA 1452
QY 535 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly--- 553
DB 1453 AGCGGCCACGCTGTGGTGTGCGGCATGTGTGACGATCTCAAGGGGTCTTCGACTCC 1512
QY 554 -----LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAsp 570
DB 1513 AACTTCCAGATCGATCAGGTGTTTCGAG-----GTCGACAAAGATGAGGAC 1557
QY 571 GlyLeuLeuLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLys 590
DB 1558 AAGCTTGTAAAGTCGCTTTTCAAGGGAACCT---CTCACTGTTCGCGGTGAGTGAACAAG 1614
QY 591 LeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGly 610
DB 1615 CTGCGCAGCAATATTCGATCGCGGGAACATGCGAGGTGTCTACTACTTCTCTGACCAG 1674
QY 611 IleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMet 630
DB 1675 TTCGAGTCAATTTCTGCTTGTGAGCAGTTTTCGATTCGAAATCTTGAAGAGCAGAGTCTG 1734
QY 631 ThrPheAlaGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
DB 1735 ACGTATGGCGAGAACTTCTTCTTCAACTTCGCGAAGTTTGTGAGTAACATCAATCCAGATC 1794
RESULT 9
DB7657 2035 bp mRNA linear PLN 07-JUL-1998
LOCUS Corallina pilulifera mRNA for vanadium-dependent bromoperoxidase 1,
DEFINITION complete cds.
ACCESSION DB7657.1 GI:3293049
VERSION DB7657
KEYWORDS vanadium-dependent bromoperoxidase 1.
SOURCE Corallina pilulifera
ORGANISM Corallina pilulifera
Eukaryota; Rhodophyta; Florideophyceae; Corallinales;
Corallinaceae; Corallinoideae; Corallina.
REFERENCE 1 (bases 1 to 2035)
AUTHORS Shimomishi,M., Kuwamoto,S., Inoue,H., Wever,R., Ohshiro,T.,
Izumi,Y. and Tanabe,T.
TITLE Cloning and expression of the gene for a vanadium-dependent
bromoperoxidase from a marine macro-alga, Corallina pilulifera
JOURNAL FEBS Lett. 428 (1-2), 105-110 (1998)
MEDLINE 98307393
PUBMED 9645486
REFERENCE 2 (bases 1 to 2035)
AUTHORS Tanabe,T.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1996) Tadaashi Tanabe, National Cardiovascular
Center Research Institute, Department of Pharmacology;
Fujishiro-dai 5-7-1, Suita, Osaka 565, Japan
(E-mail:tanabeori.ncvc.go.jp, Tel:06-833-5012, Fax:06-872-8092)
FEATURES
Location/Qualifiers
1..2035
/organism="Corallina pilulifera"
/mol_type="mRNA"
/db_xref="taxon:78447"
/clone="BP01"
128..1924
CDS

```

/codon_start=1
/product="vanadium-dependent bromoperoxidase 1"
/protein_id="BA331261.1"
/db_xref="GI:3293050"
/translation="MGIPADNLQSRKAKASFDTRVAAEALNLRGVVPSFANGBELLYR
NPDDNTDPSFIASFYTKGLPHDDNGAIIDPDDFLAFVRAINSDEKEIADLTIGPARD
PETGLPTWRDLANSLELEVRGWENSAGLTFDLEGDAQSIAMPAPVLTSPDLVAE
IAELYLMALGRETEFEFPDSPKNAEVIQPAIDOLNGLEWENTPAKLGDPPAEIRRRRG
EVTVGNLFRGLPGSVGPYLSQYIIVGSKQIGSATVGNKTLVSPNAADEFDGEIAYG
SITISQVRITPERGFMIDKLVFDQDANDFRGFSEYEPGARLIRTLKDLAIWVHF
DALYEAFLNACLILANGVFPDNLFPQDDKLDNDQDFVNFQSAHLVLSVTEVTRA
LKAVYQKQENIRLPERPTEATGLISVYNKIAPQGESIFPEVDLAVSELGDILEKARIS
NRKONTADGDDPDPSPFLPLPMFAEGSPHPSPSGHVVAGACVILKRAFFDSGIEI
DOYFEVDKQSDKLKVSFKCTLVAGELNKLADNIALGRNMAGVHVFSDQFESILLGE
QVAGILIEQSLYGENFPFNLKPDGTIQT"

```

ORIGIN

Alignment Scores:

```

Pred. No.: 1.52e-36      Length: 2035
Score: 626.00           Matches: 194
Percent Similarity: 43.77%      Conservative: 73
Best Local Similarity: 31.80%    Mismatches: 239
Query Match: 17.74%           Indels: 104
DB: 8                      Gaps: 20

```

US-10-691-383-2 (1-676) x D87657 (1-2035)

```

QY 135 ProThrAlaProAsnArgArgAspAsnVal-----AlaPheAlaSerArg 149
DB 122 CCTACATGGGTATTCAGCTGACACACCTCAAAAGTCGCGCCAAAGCTTCATTCGATACG 181
QY 150 ArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIleThrAsnGly 169
DB 182 CGTGTAGCTGGCGCGAGCTTGCACTCAACCGCGAGTTTACCATCGTTTGGCAATGGG 241
QY 170 Gln-----ThrAsp-----LeuAlaThrMet 176
DB 242 GAAGAGCTTCTTACCGCAACCCGACCTGACACACTGATCCGAGCTTCATCGCTAGC 301
QY 177 PheHisLysSerLeuProHisAspGluLeuGlyGln---ValThrAlaAspAspPheAla 195
DB 302 TTCACAAAGGGTCTTCGCGATGACGCAATGGCGCTATTATCGACCCCGACGACTTCTTG 361
QY 196 IleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspVal----- 212
DB 362 GCCTTCGTCGTCATCAATAGTGGCGATGAAGAAGAGATGCCGACCTCACATTCGGG 421
QY 213 ProAlaGlyAspProAlaGlyArgLeu----- 221
DB 422 CCAGCTCGACCGGAGACTGGCTTACCAATCTGGCGCTCGGATCTGGCGAAATTCCTC 481
QY 222 -----ValAsnProThrAlaAlaPheAlaIleAspIleSerGly 234
DB 482 GAGCTCGAAGTGGAGATGGGAGACAGCTCTGCGCGTCTCACCTTCGACCTGGAGGGC 541
QY 235 ProAlaPheSerAlaThrIleProProValProThrLeuSerSerProGluLeuAla 254
DB 542 CCGACCGCGAGTCGATTGCCATGCCACCGCGCTGTGCTCAGACGCCCTGAGCTCGTC 601
QY 255 AlaGlnLeuAlaGluLeuIleuYrTrpMetAlaLeuAlaArgAspValProPhe----- 271
DB 602 GCCGAGATAGCAGAGCTGTACTCGATGGCGCTTGGACGCGAAATTCAGATTTCAGCGAGTTT 661
QY 272 -----MetGlnTyrGlyThrAspGluIleThrThrAla 283
DB 662 GATTCCCAAGAACCGCAGAGTATATTCAGTTTGCTATTGATCATCGTTAAACGGGCTGGAG 721
QY 284 AlaAlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAlaValSerIleGlySer 303
DB 722 TGGTTCAACACACCGGCAAGCTCGGAGATCCGCTCGGAAATCCGTCGCGCGCGGT 781
QY 304 AspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThr 323
DB 323

```

```

DB 782 GAGGTGACTGTT-----GGAAACTTGTTCGCGGTATTCTTCCAGGCTCTGAGGTC 832
QY 324 GlyProPheValSerGlnLeuLeu-----ValaAsnSerPheThrIle 337
DB 833 GGCCCGTACTCAGCCAGTACATCATCGTTGGTAGCAAGCAGATTTGGCTCAGCAGAGTT 892
QY 337 ----- 337
DB 893 GGTAAACAAACTCTCGTGAGCCCAATGCTGCTGATGAGTTTGTGTTGAAATCGCCTAC 952
QY 338 AspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetVal 357
DB 953 GGAAGCATCACCATTACCCAGCGTGTGCGTATCCGCCACGCTGAGCGCGACTTCATGACC 1012
QY 358 AspPheAspGluTrpLeuAsnIleGlnAsnGlyGlyProAlaGlyProGluGluLeu 377
DB 1013 GACTTGAAGGTATTCTTTCAGCTCCAGGACGCTCGGAGCTTCCGAGGCTTTGAGTCGTAT 1072
QY 378 AspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAsp 397
DB 1073 GAGCCGGGAGCAGCGCTCATCCGAGCGATCCGCGATCTTCGAGCGTGGTGCATTTGAC 1132
QY 398 AsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSer 417
DB 1133 GCACGTGACGAGGCTACCTCAATGCGTGCCTAATTCTGTGGCGAACGCGTCCGCTTC 1192
QY 418 ArgProGlyIleAsnGlyProPheIleAspSerAspArg-----GlnAlaGlyPhe 434
DB 1193 GATCCC-----AACCTTCGTTCCAGGAGGACAGCAAGCTCGATACACGAGGAGTGT 1246
QY 435 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 453
DB 1247 GTCAACTTCGATCCGACACGACGCTGCTGAGTCTGAGTGAAGTGGCTACGCGCGCTTG 1306
QY 454 ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 473
DB 1307 AAGCGGTACGCTACGAGAGTTTAACTCATCTGCTGCGCTCGCGCTGAGGCTACCGGT 1366
QY 474 GlyThrLeuHis---AsnThrIleAla-----GlyAspLeuAspAlaAspPhe 488
DB 1367 GGTCTGATTAGCGTTTACAAAATCGCACCCGACAGGCGGAGAGCAATTTCCCTGAGGTT 1426
QY 489 AspIleSerLeuLeuAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 508
DB 1427 GATCTTGCTGTTAGAGAGCTTGAGATATCTTCGAGAAAGCTGAAATAGCAATAGGAAG 1486
QY 509 GlnAsn-----ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 523
DB 1487 CAGAACATAGCTGACGAGATCCTGACCCCTGATCTTCATTCCTGTTCGCGATGGCAATC 1546
QY 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
DB 1547 GCCGAGGCGAGCCCATTCATCTCTACGGAAGCGGCCACGCTGTGTTGCTGGCGCA 1606
QY 544 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsn 563
DB 1607 TGTGTGACGATCTCGAAGCGGTTTC-----TTCGACTCCGCGCATCGAGATC-----GAT 1654
QY 564 ProValPheProSerAsp-----AspGlyLeuGluLeuLeuAsnPheGluGlyAla 580
DB 1655 CAGGTGTTCGAGGTGCAACAAAGATGAGGACAAGCTTGTGAAGTCGTCTTCAAGGAACT 1714
QY 581 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 600
DB 1715 ---CTCACTCTTCGCGTGAATTGAACACAGCTCCCGACAATATTCGATTCGGCGGTAAAC 1771
QY 601 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 620
DB 1772 ATGCAGGTCTTCACTACTTCTCTGACCACTTCGAGTCACTTCTGCTCGGTGAGCAGGTT 1831
QY 621 ThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPhe 640
DB 1832 GCATTTGGAATCTTGAAGAGCAAGTCTGACGTATGCGCGAAGCACTTCTTCTTCAACTTG 1891

```

```

Qy 641 ArgLeuPheThrGlyGluValIleLeu 650
Db 1892 CCGAAGTTTGATCGAAGTACAATCCAGATC 1921

RESULT 10
LOCUS E17199 1794 bp DNA linear PAT 28-JUL-1999
DEFINITION Corallina pilulifera mRNA for haloperoxidase.
ACCESSION E17199
VERSION E17199.1 GI:5711882
KEYWORDS JP 1998248581-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1794)
AUTHORS Izumi, Y. and Tanabe, T.
TITLE NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL Patent: JP 1998248581-A 1 22-SEP-1998;
COMMENT OTSUKA PHARMACEUT CO LTD
PN JP 1998248581-A/1
PD 22-SEP-1998
PF 06-MAR-1997 JP 1997070539
PI IZUMI YOSHIKAZU, TANABE TADASHI
PC C12N15/09, C12N9/08, (C12N15/09, C12R1:89), (C12N9/08, C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..1794
FH /organism="Corallina pilulifera" FT
FT /clone="BP01"
FT mat_peptide 1..1794
FT Location/Qualifiers
FT 1..1794
FT /organism="unidentified"
FT /mol_type="genomic DNA"
FT /db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 2,12e-36 Length: 1794
Score: 623.00 Matches: 191
Percent Similarity: 44.41% Conservative: 75
Best Local Similarity: 31.89% Mismatches: 233
Query Match: 17.66% Indels: 100
DB: 6 Gaps: 20

US-10-691-383-2 (1-676) x E17199 (1-1794)
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThr 160
Db 31 CGGCCAAGGCTTCATTGATACGCGGTAGCTCGCGCGAGCTTGCACTCAAC---CGC 87
Qy 161 GlyThrValCysGlnIleThrAsnGlyGlu 170
Db 88 GGAGTTGTACCATCGTTGGCAATGGGAGAGAGCTTCTTACCGCAACCCGGACCTGAC 147
Qy 171 ---ThrAsp-----LeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGly 187
Db 148 AACACTGATCCGAGCTTCATCGTAGCTTCACAAAGGGTCTTCGCGCATGACGACAAATGGC 207
Qy 188 Gln---ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
Db 208 GCTATTATACGCCCGACGACTTCTTGGCCCTTCGTCGCAATCAATAGTGGCGATGAA 267
Qy 207 SerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
Db 268 AAGGAGATCCCGGACCTTCATATGGGGCCAGCTCGCGACCCGGAGACTGCTTACCAATC 327
Qy 222 -----ValAsnProThr 225

```

```

Db 328 TGGCGCTCGGATCTGGCGAAATTTCTCTCGAGCTCGAAGTGCAGGATGGGAGACAGCTCT 387
Qy 226 AlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProProVal 245
Db 388 GCCGGTCTCACCTTCGACCTGGAGGCGCGGACCGCGATTCGATTCGATTCGACCGCGG 447
Qy 246 ProThrLeuSerSerProGluLeuAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeu 265
Db 448 CCTGTGCTCAGAGCCCTGAGCTCGTCCGCGAGATAGCAGAGCTGTACTCTGATGGCGTT 507
Qy 266 AlaArgAspValProPhe-----MetGlnTyr 274
Db 508 GGACCGAAATCGAGTTTCAGCGAGTTTCCCAAGAACGCGAGAGTATATTTCAGTTT 567
Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 568 GCTATTGATCAGCTTAACGGGCTGAGTGGTTCAACACACCGGCAAGAGCTCGAGATCCG 627
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspProPheSerGlnLeuPhe 314
Db 628 CCTCGGAAATCCGTCGCCGCGGTGAGGTGACTGTT-----GGAAACTTGTTC 678
Qy 315 ArgAlaThrPheValGlyValGluThrGlyProPheValSerGlnLeuLeu----- 331
Db 679 CGCGGTATTCTTCAGGCTCTGAGTCCGCCCGCTACCTCAGCCAGTACATCATCGTTGGT 738
Qy 332 -----ValAsnSerPheThrIle----- 337
Db 739 AGCAGCAGATTTGGCTCAGCGACAGTTGGTTAAACAAACTCTCGTGAGCCCCAATGCTGCT 798
Qy 338 -----AspAlaIleThrValGluProLysGlnGluThr 348
Db 799 GATGAGTTTGATGGTGAATCGCCTACGGAAGCATCACCATTAGCCAGCGTGCATC 858
Qy 349 PheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeuAsnIleGlnAsnGly 368
Db 859 GCCACGCTCGGACGCGACTTCATGACGCGACTTGAAGGTATTCTCTTGAAGTCCAGGCGCT 918
Qy 369 GlyProProAlaGlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArg 388
Db 919 CGGAGCTTCGAGCTTTGAGTCGTATGCGCGGGAGCAGCCCTCATCCGAGCATCCGC 978
Qy 389 AspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaIleThrArgGlySerLeu 408
Db 979 GATCTCGGAGCTGGGTGCGACTTTGACGCGACTGACGAGGCTACCTCAATGCGTGCCTA 1038
Qy 409 IleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer 428
Db 1039 ATTCTGTTGGCAACGCGCTGCGCTTCGATCCC-----AACCTTTCGTTCCAGCAGGAG 1092
Qy 429 AspArg-----GlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeu 445
Db 1093 GACAACTCGATTAACGAGGAGCGTGTGTGAACTTCGAGTCCGCGACAGCTGCTGAGTCTG 1152
Qy 446 Ile---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHis 464
Db 1153 GTGACTGAAGTGCTACGCGCGGTTGAGGCGGTACGGTACCGAGAGTTTAACTTCAT 1212
Qy 465 ArgPheAlaArgProGluAlaLeuGlyGly---ThrLeuHisAsnThrIleAla----- 481
Db 1213 CGTCGCTGCGCCCTGAGGCTACCGGTGGTCTGATTATCGTTTAAACAAATCGCACCGCAG 1272
Qy 482 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeu 499
Db 1273 AAGGGCAGAGCATTTTCCTCGAGTTGCTTCTGCTGTGAAGAGCTTGGAGATATCTTG 1332
Qy 500 LysArgValAlaGluIleAsnAlaAlaGlnAsn-----ProAsnAsnGlu 514
Db 1333 GAGAAAGCTGAATTAGCAATAGGACGACGACATAGCTGACGAGAGTCTCTGACCTGAT 1392
Qy 515 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 534

```

Db 1393 CTTTCATTCCTGTTGCCGATGCGATTCGCCGAGCGCAGCCCATTCATCCGTCCTACCGA 1452

Qy 535 SerGlyHisAlaThrGlnAnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 554
|||||
|||||

Db 1453 AGCGGCACCGCTGGTGTGGCGCATGTGTACGATCCTGAAGCGCTTC-----TTC 1506

Qy 555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAsp-----AspGly 571
|||
|||

Db 1507 GATCCGCCATCAGATC-----GATCAGTGTTCGAGTGCACAAAGATGAGGACAG 1560

Qy 572 LeuGluLeuIleAsnGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeu 591
|||
|||

Db 1561 CTTGTGAAGTCGTCTTTCAAGGAACT---CTCACTGTTCCGCGTGAATTTGAACAAGCTC 1617

Qy 592 AlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle 611
|||
|||

Db 1618 GCCGACAATATTCGATCGCGCGTAACATCGCAGGTGTTCTCACTTCTCTGACCAAGTTC 1677

Qy 612 GlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThr 631
|||
|||

Db 1678 GATCACTTCTGTCGGTGAGCAGGTGGCATTTGGATCTTTGAAGAGCAAGCTCTGACG 1737

Qy 632 PheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
|||
|||

Db 1738 TATGGCGAGAAGTCTCTTCAACTTCCGGAAGTTTGATGGAAGTCTACAAATCCAGATC 1794

RESULT 11

Db 7658

LOCUS 2029 bp mRNA linear PLN 07-JUL-1998

DEFINITION Corallina pilulifera mRNA for vanadium-dependent bromoperoxidase 2, complete cds.

ACCESSION D87658

VERSION D87658.1 GI:3293051

KEYWORDS vanadium-dependent bromoperoxidase 2.

SOURCE Corallina pilulifera

ORGANISM Corallina pilulifera
Eukaryota; Rhodophyta; Florideophyceae; Corallinales;
Corallinaceae; Corallinoideae; Corallina.
1 (bases 1 to 2029)
Shimomishi, M., Kuwamoto, S., Inoue, H., Wever, R., Ohshiro, T., Izumi, Y. and Tanabe, T.
Cloning and expression of the gene for a vanadium-dependent bromoperoxidase from a marine macro-alga, Corallina pilulifera
FEBS Lett. 428 (1-2), 105-110 (1998)

JOURNAL 98307393

MEDLINE 9645486

REFERENCE 2 (bases 1 to 2029)
Tanabe, T.
Direct Submission
Submitted (02-SEP-1996) Tadaashi Tanabe, National Cardiovascular Center Research Institute, Department of Pharmacology,
Fujishiro-dai 5-7-1, Suita, Osaka 565, Japan
(E-mail: tanabe@ri.ncvc.go.jp, Tel: 06-833-5012, Fax: 06-872-8092)
Location/Qualifiers
1. 2029
/organism="Corallina pilulifera"
/mol_type="mRNA"
/db_xref="taxon:78447"
/clone="BPO2"
119..1912
/codon_start=1
/product="vanadium-dependent bromoperoxidase 2"
/protein_id="BA31262.1"
/db_xref="GI:3293052"
/translation="MGIPADNLOSRAKASFDTRVAAAEALALARGVVPSPFANGBELLYR
NCTGDPSTFIASGTGPDHDDNGAIIDPDDFLAFRAINSDEKEIADLTGLPARDPE
TGLPIWRSDLNLELEVRGWENSGASLTFDLEGPOAQSVAMPAPVLPMSDELIAEWA
EYLWALGRDIEFSESPKNAAPIRSAIERLNGLEWNTPAKLGPPAPRIERRGEV
TVGNLRGILPSEVGPYLSQYIIVGSKQIGSATVGNKTVSPNADEPDGEIAYGSI
TISCRVATPGDFTDLKFLVDVQDAGDFGESYEPGARLIRTLRLDVAWFDK
LISAYLNACILILANGVFPDNLFPQOEDKLDNHDVFNFGSAHVLSTVEVATRAIK
AVRYOKFNIRHRLPRTGGLIIVNKKSLAGSDIIFPEVSELSELSSILDDVAESN
EKQNRADGIVSPDKSFLPPMAFAGEGSPFHPSPGSHAVAGACVTLKAPFDANFQID

FEATURES
source

CDS

ORIGIN

Alignment Scores: 6.35e-36 Length: 2029
Pred. No.: 617.50 Matches: 201
Score: 41.12% Conservative: 77
Percent Similarity: 41.12% Mismatches: 229
Best Local Similarity: 29.73% Indels: 169
Query Match: 17.50% Gaps: 24
DB: 8

US-10-691-383-2 (1-676) x D87658 (1-2029)

Qy 67 ProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArg-----Val 84
|||||
|||||

Db 113 CCACAAATGGGTATTCAGCTGCAACCTCAAAGTCGCGCCCAAGGCTTCATTCGATACG 172

Qy 85 ArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluVal 104
|||
|||

Db 173 CGTGTAGCTCGGCC-----GAGCTGGCA 196

Qy 105 ThrAlaSerGlyValPheProGluGluHisLysHisThrGlyGluGlyArgHisLeuGln 124
|||
|||

Db 197 CTGCGCGGTGAGTTGTGCCA----- 217

Qy 125 ThrCysThrAsnSerAspAlaLeuAspProThrAlaProAsnArgArgAspAsnVal 144

Db 217 ----- 217

Qy 145 AlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThrGlyThrValCys 164
|||
|||

Db 218 TCGTTTGCATAATCGAGGAGCTTCTTACCGCAAC-----TGT 256

Qy 165 GlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAsp 184
|||
|||

Db 257 GAG-----ACCGCGATCCAAGCTTCATTTGCTAGTTTCAAAAGGCTTCCACATGAC 310

Qy 185 GluLeuGlyGln---ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsn 203
|||
|||

Db 311 GACAATGGCGCTATTATTATTCGCGAGATTTCTTGGCCTTCGTTCTGCGGATCAATAGT 370

Qy 204 GlyAspPheSerIleCysGluAspVal-----ProAlaGlyAspProAlaGlyArg 220
|||
|||

Db 371 GGAGATGAGAGGAGATCGCGGACCTTACATTCGCGGCCAGCTCGCGACCTCGAGACTGC 430

Qy 221 Leu-----Val 222

Db 431 TTACCAATCTGGCGCTCAGATCTGGCGAATTTCTCGAGCTCGAAGTCGAGGATGGAA 490

Qy 223 AsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIle 242
|||
|||

Db 491 AACAGCTCTCGCGTCTTACTTCGACCTGGAGGCGCGGACGCGGAGTGGTGGCCATG 550

Qy 243 ProProValProThrLeuSerSerProGluLeuAlaGlnLeuAlaGluLeuTyrTrp 262
|||
|||

Db 551 CCACCGCGCTGTGCTCATGAGCCCTGAGCTCATCGCGAGATGGCAGAACTGTACTCTG 610

Qy 263 MetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGlu-----Ile 279
|||
|||

Db 611 ATGCGCTTGGAGCGCAGATCGAGTTTAGCGAGTTTCGAGTTCGCCAAGAACGACGACTTTT 670

Qy 280 ThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAla--- 298
|||
|||

Db 671 ATTAGGTCTGCTATTGAACCGGCTTAACGGGCTGGAGTGGTTC---AACACACCGGCAAG 727

Qy 299 -----ValSerIleGlySerAsp 304
|||
|||

Db 728 CTTGGTATCCGCTCGGAAATCCGTCGCGCTCGCGGTGAGGTGACCGTCGGAAC--- 784

Qy 305 GlyThrValAspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGly 324
|||
|||

Db 785 -----TTGTTCCGCGGTATTCTTCCCGGTTCCTCCCGGTTCGAGGTCGCG 820

```
Qy 325 ProPheValSerGlnLeuLeu-----ValAsnSerPheThrIle--- 337
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 CCGTATCTCAGCCAGTACATCATCTGTTGGTAGCAAGCAGATTGGCTCTCGCAGCGTTGGC 880
338 -----Asp 338
881 AACAAAACTTTCGTGAGCCCAATGCGCGTGATGAATTTGATGTGAAATCGCCTACGGA 940
Qy 339 AlaIleThrValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAsp 358
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 941 AGCATCACCATTGCCAGCGTGGCTATCGCCAGCGCTGGCGCGACTTCATGACTGAT 1000
359 PheAspGluTrpLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAsp 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1001 TTGAAGGTATTCCTTGAGTCCAGTCCAGACGGTGCAGACTTCGAGGATTTGAGTCGTATGAG 1060
379 GluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsn 398
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1061 CCGGAGCAGCGCTCATCTCGGACGATCGCGATCTTCGACGTGGTGCACTTTTGACGCA 1120
399 IleAsnThrGluAlaTyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArg 418
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1121 CTGTACGAGGCTACCTCAATGCGTGTGATCTGTTGGCAACGGCGTCCGTTTCGAT 1180
419 ProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGly-----PheVal 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1181 CCC-----AACCTTCCTTCCAGCAGGAGGACAACTCGACCAACCATGATGTGTTGTG 1234
436 AsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArg 454
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1235 AACTTCGGATCCGACACAGTCTGAGTCTGTGATCTGAAGTGGCTACGCGCGCTTGAAG 1294
455 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1295 CGGTGACGGTATCAGAAGTTTAACATTCATCGTCGCGCTGAGGCTTACTGGTGGT 1354
475 ThrLeu-----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIle 490
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1355 CTGATTAGCTGTTAACAGAAATCTTCTTGGCGGTAGT-----GATATA 1399
491 SerLeuLeuGluAsnAspGluLeuLeuLysArg-----ValAla 503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1400 ATATTCCTCGAAGTTAGTAGTACTGCTGAGGAGCTCTCATCAATTCGTGATGAGCTTGTCT 1459
504 GluIleAsnAlaAlaGlnAsnProAsnAsnGluVal-----ThrTyrLeu 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1460 GAGAGCATGAAACAGAACAGCGGCTGACGGGATTGTAGCCCGCGGATAAATCATTTCTG 1519
519 LeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAla 538
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1520 TTGCGGATGGCATTTGCGAGGCGAGCCCATTCATCGTCTATGGAAGTGGCCACGCT 1579
539 ThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly-----Leu 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1580 GTGGTTGTCGGCGCATGTGTGACAATCTCAAGCGTCTTCGACGCCCACTTCCAGATC 1639
555 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1640 GATAAGGTGTTGCGAG-----GTCGACACTGATGAGGACAAGCTTGTGAAG 1684
575 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 594
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1685 TCGTCTTTCAGGGAAGCT---CTCACTGTTGCCGTGATGTAACAGCTCGCCGACAT 1741
595 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1742 GTTCGGATCGGGCGGAACATCGCGGTGTTCACTACTTCTCTGACCAAGTTTCGAGTCACTG 1801
615 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGlu 634
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1802 CTACTCGGTGAACAGATTGGGATTTGGAATCTTGGAGGAGCAGACCTTGACGTATGGCGAG 1861
```

```
Qy 635 GluAlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 650
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1862 AACTTCTTCTCAATTTGCGAAATTCGACGGAAGTACAAATCCAGATT 1909

RESULT 12
E17200
LOCUS E17200 1791 bp. DNA linear PAT 28-JUL-1999
DEFINITION Corallina pilulifera mRNA for haloperoxidase.
ACCESSION E17200
VERSION E17200.1 GI:5711883
KEYWORDS JP 1998248581-A/2.
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1791)
AUTHORS Izumi, Y. and Tanabe, T.
TITLE NEW HALOPEROXIDASE GENE AND ITS UTILIZATION
JOURNAL Patent: JP 1998248581-A 2 22-SEP-1998;
OSUTSUKA PHARMACEUT CO LTD
COMMENT Corallina pilulifera
PN JP 1998248581-A/2
PD 22-SEP-1998
PF 06-MAR-1997 JP 1997070539
PI IZUMI YOSHIKAZU, TANABE TADASHI
PC C12N15/09,C12N9/08,(C12N15/09,C12R1:89),(C12N9/08,C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1791 /organism='Corallina pilulifera' FT
FT /clone='BPO2' /mat_peptide 1..1791
FT /product='haloperoxidase'.
FEATURES
source Location/Qualifiers
1..1791
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 8.9e-36 Length: 1791
Score: 614.50 Matches: 193
Percent Similarity: 43.04% Conservative: 70
Best Local Similarity: 31.59% Mismatches: 223
Query Match: 17.42% Indels: 125
DB: 6 Gaps: 21

US-10-691-383-2 (1-676) x E17200 (1-1791)
Qy 141 ArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArgGluArgAspGlyThr 160
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 31 CGCGCCAGGCTTCATTGATGACGGTGTAGTCTCGCGCGGAGCTGCGCACTC---GCCCGT 87
161 GlyThrValCysGlnIleThrAsnGlyGlu----- 170
88 GGAGTTGTGCCATCGTTTGCAAATGGAGAGAGCTTCTTACCGCAACTGTGTGACACCGC 147
171 ---ThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeuGlyGln--- 188
148 GATCCAAAGCTTCATTGCTAGCTTCCACAAAGGCTTCTCCACATGACCAATGGCGCTATT 207
189 ValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIle 208
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 208 ATTGATCCCGACGATTTCCTTGGCCCTTGTGCGGATCAATGAGTGGAGATGAGAAGAG 267
209 CysGluAspVal-----ProAlaGlyAspProAlaGlyArgLeu----- 221
268 ATCGCCGACCTTACATTGGGGCCAGCTCGGACCTCGAGACTGGCTTACCAATCTGGCGC 327
222 -----ValAsnProThrAlaAla 227
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 328 TCAGATCGCGAATCTCTCGAGCTCGAAGTGGAGGATGGGAAAAACAGCTCTGCCGGT 387
Qy 228 PheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIleProValProThr 247
Db 388 CTTACCTTCGACCTGGAGGCGCGAGCGGAGTGGTTCGATGCCACCGCGCCCTGTG 447
Qy 248 LeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArg 267
Db 448 CTCATGAGCCTGAGCTCATCGCCGAGATGGCAGACTGTACCTGATGGCGCTGGAGCG 507
Qy 268 AspValProPheMetGlnTyrGlyThrAspGlu-----IleThrThrThrAlaAla 284
Db 508 GACATCGAGTTTACGAGTTTCGAGTCCCAAGAACGACGCTTTTATTAGTCTGCTATT 567
Qy 285 AlaAsnLeuAlaGlyMetGlyPheProAsnLeuAspAla----- 298
Db 568 GAACGGCTTAACGGCTGGAGTGTTC---AACACCGCGCAAGCTTGTGTGATCCGCT 624
Qy 299 -----ValSerIleGlySerAspGlyThrValAspPro 309
Db 625 GCGGAATTCGTCGCCGTCGGGTGAGTGCACGTCGGAAC----- 666
Qy 310 PheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheValSerGln 329
Db 667 -----TTGTTCCGCGGTATTCTTCCCGGTCTCGAGTGGCGCGGTATCTCAGCCAG 717
Qy 330 LeuLeu-----ValAsnSerPheThrIle----- 337
Db 718 TACATCATGTTGTTAGCAAGCAGATTGGCTCTCGGACGTTGGCAACAAACTTTCGTG 777
Qy 338 -----AspAlaIleThrValGlu 343
Db 778 AGCCCCAATGTCGCGATGAATTTGATGTTCAATCCCTACCGAAGCATCACCAATTAGC 837
Qy 344 ProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGluTrpLeu 363
Db 838 CAGCGTGGCGTATCGCCACCGCTGGCGGAGCTTCATGACTGATTTGAAGGTATTCCTT 897
Qy 364 AsnIleGlnAsnGlyGlyProProAlaGlyProGluLeuAspGluLeuAspGlyPhe 383
Db 898 GAGCTCAACAGCGTGCAGACTTCCGAGGATTTGAGTGTGATGAGCGGAGCAGCGCTC 957
Qy 384 IleArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAla 403
Db 958 ATTCGAGCATCCCGATCTTCGACGTGGGTGCATTTGACGCACTGTACGAGCCCTAC 1017
Qy 404 TyrArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGly 423
Db 1018 CTCAAATCGTCCTGATTCGTGGCGAAGCGCGTGCCTTCGATCCC-----AACCTT 1071
Qy 424 ProPheIleAspSerAspArgGlnAlaGly-----PheValAsnPheGlyThrSer 440
Db 1072 CCTTTCAGCAGGAGGACAACTCGAACCATGATGTGTTTGAACCTTCGGATCCGCA 1131
Qy 441 HisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 459
Db 1132 CACGTGTGAGTCTGGTGAAGTGAAGTACGCGCGCTGTGAAGCGGTACGGTATCAG 1191
Qy 460 LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu----- 476
Db 1192 AAGTTTAACATTCATCGTCGCGCTCGCGCTGAGGCTACTGTGTCTGATTATCGTTAAC 1251
Qy 477 ---HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495
Db 1252 AAGAAATCTTCTTCCTGCGGGTAGT-----GATATAATATTCCCTGGAAGTT 1296
Qy 496 AspGluLeuLeuLysArg-----ValAlaGluIleAsnAlaAla 508
Db 1297 AGTGAATCTGTCGAGGAGCTCTCATATTCTGATGACGTTCGTGAGCAATGAAAA 1356
Qy 509 GlnAsnProAsnAsnGluVal-----ThrTyrLeuLeuProGlnAlaIle 523
Db 1357 CAGAACAGGGCTGACGGGATTGTGAGCCCGGATTAATCATTTCTTCTGTTGCCGATGCGCATTT 1416

Qy 524 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 543
Db 1417 GCGAGGGCAGCCATTCATCGTCTATGAAGTGGCCACCGCTGTGGTGGCGCA 1476
Qy 544 PheAlaThrValLeuLysAlaLeuIleGly-----LeuAspArgGlyGlyGlu 559
Db 1477 TGTGTGACATCTCAAGCGGTTCTTCGACGCCAACATTCAGATCGATAAGGTGTCGAG 1536
Qy 560 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly 579
Db 1537 -----GTCGACACTGATGAGCAAGCTTCTGAAGTCTCTTCAAGGGA 1591
Qy 580 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 599
Db 1582 ACT---CTCACTGTTCGGTGAATGAACAGCTGCCACCAATGTTGGATCGGCGG 1638
Qy 600 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 619
Db 1639 AACATGGCGGTGTTCACTACTTCTCTGACCAGTTCGAGTCACTGCTACTCGTGAACAG 1698
Qy 620 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGlu 639
Db 1699 ATTGCGATTGGAATCTTGGAGGACGAGCTGACGTATGCGAGAACTTCTTCTTCAAT 1758
Qy 640 PheArgLeuPheThrGlyGluValIleLysLeu 650
Db 1759 TTCCCGAATTCAGCGAACTACATCCAGATT 1791

RESULT 13

LOCUS BX294153 293350 bp DNA linear BCT 11-JUL-2003
DEFINITION Pirellula sp. strain 1 complete genome; segment 21/24.
ACCESSION BX294153 BX119912
VERSION BX294153.1 GI:32447383
KEYWORDS complete genome.
SOURCE Pirellula sp. 1
ORGANISM Pirellula sp. 1
Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
Planctomycetaceae; Pirellula.
REFERENCE 1 (bases 1 to 293350)
Gloeckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T.,
Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R.,
Schlesner, H., Amann, R. and Reinhardt, R.
Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1
Proc. Natl. Acad. Sci. U.S.A. 100 (14), 8298-8303 (2003)
MEDLINE 22735913
PUBMED 12835416
REFERENCE 2 (bases 1 to 293350)
Kube, M., Borzym, K., Heitmann, K., Klages, S., Marquardt, I.,
Lehrack, S., Beck, A., Pawlik, R., Reinhardt, R., Gloeckner, F.O.,
Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Rabus, R.,
Schlesner, H. and Amann, R.
Direct Submission
Submitted (21-JAN-2003) Max Planck Institute for Molecular
Genetics, ProScience Ihnestrasse 73, D-14195 Berlin, Germany Max
Planck Institute for Marine Microbiology Celsiusstrasse 1, D-28359
Bremen, Germany
TITLE
JOURNAL
COMMENT
This project was carried out by
*Max Planck Institute for Molecular Genetics, Berlin, Germany; *Max
Planck Institute for Marine Microbiology, Bremen, Germany; in the
framework of the REGX-project, <http://www.regx.de> -----
Genome Center
Center: Max Planck Institute for Molecular Genetics
Center code: MPIMG
----- Summary Statistics
Sequencing vector: pUC19; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 7142841 bases at least Q40
Consensus quality: 7145138 bases at least Q30
Consensus quality: 7145484 bases at least Q20

Quality coverage: 8.03

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid sequence; assembly was additionally confirmed by long range PCR and cosmid end sequences.

See <http://www.micro-genomes.mpg.de/pirellula/> for more information including minimal tiling path from a set of 220 cosmids out of 908. See the misc_feature tag below for the boundaries of the MTP cosmids. ----- Annotation

Center: Max Planck Institute for Marine Microbiology
Celliusstrasse 1, D-28359 Bremen, Germany.

Center Code: MPIWM

Email: fogempi-bremen.de

Phone: +49 (0)421 2028 938 Fax: +49 (0)421 2028 580

Three different programs (Glimmer, Critica and Orpheus) were used for ORF-prediction. A nonredundant list of ORFs was generated by suitable parsing of the results.

Automated annotation was done with the software package Pedant Pro (<http://www.biomax.de>). All ORF predictions and annotations were manually corrected by considering all results of the different tools applied. See <http://www.regx.de> for more information and access to supplementary information.

FEATURES source

Location/Qualifiers
1. .293350
/organism="Pirellula sp. 1"
/mol_type="genomic DNA"
/strain="1"
/db_xref="taxon:243090"
complement(93..1040)
/gene="uspA"
/locus_tag="RB11183"
complement(93..1040)
/gene="uspA"
/locus_tag="RB11183"

gene

CDS

/notes="PMID: 9371463 best DB hits: BLAST: pir:A69220;
conserved hypothetical protein MTH898 - Methanobacterium;
E=1e-07 gb:AA046412.1; AF096262.1 (AF096262) ER6 protein
[Lycopersicon; E=1e-05 gb:AA041173.1; AB006246.3
(AB006246) conserved hypothetical; E=2e-05 COG: MTH898;
COG589 Universal stress protein UspA and related; E=1e-08
PFAM: PF00582; Universal stress protein family; E=4e-11"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD78973.1"
/db_xref="GI:32447384"

gene

CDS

/translation="MRVLLAVDSPPYQQAVFASHPLRKPVDVLVSVAAPMLVD
TGMSPMPDFSGLEIFEDRSREDAIVASDLKSDQHVHVPVPIGPPTSAALLDVA
DEGADILVIGALGHSIAIERVLGSDVYVATADMTLVVRPTSEADVPDLQKIML
ALSGRDEMLTWRELKLPNVHLVDRFSYVRODLRQQASDAWAOHQEQAO
AQLDFTKLQQLGNTETHEFVSNHVGTELVAREHGGDLAVTGDSGSLTRVFL
GSTRVYLRADCSVLIIRDREDAKARQIAEQSLAST"
complement(1116..1241)
/locus_tag="RB11184"
complement(1116..1241)
/locus_tag="RB11184"
/codon_start=1
/transl_table=11
/product="hypothetical protein-signal peptide prediction"
/protein_id="CAD78974.1"
/db_xref="GI:32447385"
/translation="MFASRWTVVRAASLTASPTCATPSHLARHSQQNASWSEPKK"
complement(1231..2142)
/gene="uspA"
/locus_tag="RB11185"
complement(1231..2142)

gene

CDS

/gene="uspA"
/locus_tag="RB11185"
/notes="PMID: 8905231 best DB hits: BLAST:
Swissprot:P74148; YD88 SYN3 HYPOTHETICAL 17.3 KD PROTEIN
SL11388; E=2e-05 gb:AAG20091.1; (AE005088) Vng1898C
[Halobacterium sp. NRC-1]; E=5e-05 swissprot:Q57951;
Y511-METUA HYPOTHETICAL PROTEIN M30531 -----; E=7e-04 COG:
sl11388; COG0589 Universal stress protein UspA and
related; E=2e-06 PFAM: PF00582; Universal stress protein
family; E=0.04"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein-putative
universal stress protein"
/protein_id="CAD78975.1"
/db_xref="GI:32447386"
/translation="MNERNGPKLRLTIGISSAPARDALQALSKSTLAEECEVSLATI
VPEPLYTIDDTSMPIPOEVFSDRLRELENTOLKQOEFGERPSSCFSSQGHGPR
GLIEMSERFDADWIVIGSVGHSAFSAFRIILGSTSDYVANRSTRCLVHRPITTDHEPSK
GTFVIAISNAESTALPDWVAALRVNCEVHLVNMETHPEYAEHAKCDLTVVGDDDD
SEYRSAAWKLMTETPRLEALGMKVKFSLLESHPVGRVAVLEYANEHACDLTVVGDDDD
SLMERVWLGSVSFRVVRHANQSVLIVR"
complement(2139..3614)
/gene="sasR"
/locus_tag="RB11186"
complement(2139..3614)
/gene="sasR"
/locus_tag="RB11186"
/function="two-component signal transduction"
/notes="best DB hits: BLAST: gb:AAG45333.1; AF273214_1
(AF273214) SasR [Mycococcus xanthus]; E=3e-83
Swissprot:Q06065; ATOC ECOLI ACETOACETATE METABOLISM
REGULATOR; E=1e-81 gb:AAC28085.1; (AF100457) response
regulator [Mycococcus xanthus]; E=5e-80 COG: atcC; COG2204
AAA superfamily ATPases with N-terminal receiver; E=1e-82
PFAM: PF00072; Response regulator receiver doma; E=2.7e-32
PF00158; Sigma-54 interaction domain; E=1.1e-119"
/codon_start=1
/transl_table=11
/product="NtrC-like response regulator SasR"
/protein_id="CAD78976.1"
/db_xref="GI:32447387"
/translation="MIGITMSQTAIKILLVDDDDYRRSCGKPMERMGHTVRLAASGA
EAMSFIDREAFDVAFIDIMPMSGELMQRVHEESHDEVFLTGKGTIACVQAMQ
MGADFLTPCSDLADLEHRVQLAQGHKKENQQLKAIVHRNRPSCALIGNSLAMKE
LSMIAKVAPTKRPVLIEGSGTGKEVARSVQQSLADRPFTVINCAALPAOLVES
ELFCHQKGSFTGATAQDGLFEVADGGLFIDVSGELPPALQPKLRLVEDGSLRRIG
CHERKVKVRLIAATNRDLLEEVKAGRFREDLLYRINVLINLPLPQRSGDIELLLH
HFLPRGWSFDKEALDTMVQVHWPNRQLINVERATILADHQVITMDDLPRLGSLNP
KVLQELGKNKAKTARHLGIRHRRKLYRLLERFEQQMPAETL"
complement(3740..4012)
/locus_tag="RB11190"
complement(3740..4012)
/locus_tag="RB11190"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD78977.1"
/db_xref="GI:32447388"
/translation="MIEISHGARRQRNGNNTSPFVLKTRASITRSHRERHAMLFFV
TPPLNPTCPNSRPLRHSIRFDADVVTGLTDRFAGCFAPRPAPV"
3980..5446
/gene="hydH"
/locus_tag="RB11191"
3980..5446
/gene="hydH"
/locus_tag="RB11191"
/EC_number="2.7.3.-"
/function="two-component signal transduction"
/notes="PMID: 2666400 PMID: 11243806 best DB hits: BLAST:
gb:AAG59806.1; AF305914.1 (AF305914) HydH [Klebsiella
oxytoca]; E=2e-23 gb:AAG09417.1; AF180147_9 (AF180147)


```
misc_feature
664..665
/notes="cosmid pircos-c3h04/ cosmid pircos-blf02 joining
point"
1740..2954
/locus_tag="RB1856"
CDS
1740..2954
/locus_tag="RB1856"
/EC_number="4.2.1.24"
/function="biosynthesis of vitamins, cofactors, and
prothetic groups"
/notes="PMID: 8226669 best DB hits: BLAST: dbbj:BAA35069.1;
(AB015492) porphobilinogen synthase; E=1e-114
swissprot:P45622; HEM2 BRAJA DELTA-AMINOLEVULINIC ACID
DEHYDRATASE; E=1e-108 wisprot:Q59643; HEM2 PSEAE
DELTA-AMINOLEVULINIC ACID DEHYDRATASE; E=1e-105 COG:
PA5243; COG1113 Delta-aminolevulinic acid dehydratase;
E=1e-106 PFAM: PF00490; Delta-aminolevulinic acid
dehydratase; E=1.8e-183"
/codon_start=1
/transl_table=11
/product="delta-aminolevulinic acid dehydratase"
/protein_id="CAD72300.1"
/db_xref="GI:32396997"
/translations="MGWVGRGEIRKRPFPNVIRGSIARPGRLGTFCDAKWKHLAA
DECPKITSAPVLSILWVFLSFGSPAPRAFPATRLRRVRATOWSRRLVRETTLS
VDLILWFLVMDGSGQOPVCSLGVNRLGESEIVAARARVDLGI PAIALPAPDPKL
KSEDAEAYNSDLNCRVTKIQIDAVGDSLGVLVDALPDYSSHGQDGLVQDQVIND
EYDVLCKQAIQVAGSCDVAPSDMMWDGRIGAIRSALDGAHSGVQIISYAAKYASA
FYGFPRDAGNAAAGAAKKTKYQSPQSDEAIAEVALDLAEGADSVMVKPQMPYLD
IVARVQTFCVPTFAVQVSGEYAMLRGAADAGWLSGDAVILESLLSFKRAGADGVLY
FAADAAELLHRS"
2980..3774
/locus_tag="RB1858"
CDS
2980..3774
/locus_tag="RB1858"
/function="protein modification (glycosylation,
acylation, myristylation, palmitylation, farnesylation and
processing)"
/notes="PMID: 12093901 best DB hits: BLAST: pir:H82553;
3-demethylubiquinone-9 3-methyltransferase XF2471; E=0.027
pir:T30158; hypothetical protein C37A2.6 - Caenorhabditis
elegans; E=0.030 embi:CAA89660.1; (Z49629) ORF YJR129c
[Saccharomyces cerevisiae]; E=0.76 COG: XF2471; COG2227;
E=0.003"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD72301.1"
/db_xref="GI:32396998"
/translations="MVFDQPAHSASRSSTPMSNATVLEHETGDKTLLFSELSQT
QFMSADLSFAWHEDTLIAGQERPLAVASDPDAMLIACERDAGEGVDPFWAIT
WRAAGLDRELDVPIHNDVLRVGGCTGAGIALLRGARVLTDTGVEDPLQVRLS
LRLGLHADVQVRLGSDSLAPKPPFLLGSDVTLRLTWPELLOSAEHLTBDGOLI
LSDPQRLIAIEFSSQWVKDPEVDTEHVMEDDDPEHPIRIMVLTQG"
3822..4206
/locus_tag="RB1859"
CDS
3822..4206
/locus_tag="RB1859"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAD72302.1"
/db_xref="GI:32396999"
/translations="MSDFASRLDPNGSGESYBESSNVLRFSPKQSQRTVVPNCLRVD
GSIESREANAVNRSGVSHDVIIRGGTIGGNCSTLVKARACIALFQSIGSSVRYNLT
YAVTHAPKQSRPGSTWRRLDGG"
4261..4593
/locus_tag="RB1860"
CDS
4261..4593
/locus_tag="RB1860"
/codon_start=1
/transl_table=11
```

```
/product="hypothetical protein-signal peptide prediction"
/protein_id="CAD72303.1"
/db_xref="GI:32397000"
/translation="MSVRYNLTYSKPAQGVPPGGGWMVNIALASRLILASLCSKAS
AMPFGTTPWPGCPAKTKSKSLDPTTYLLYRLTSTPYNQLMASVRTCSLQSTLPTAFR
SILRSFAK"
gene
4530..4937
/genes="gntR"
CDS
4530..4937
/locus_tag="RB1862"
/genes="gntR"
/function="transcriptional control"
/notes="PMID: 9384377 best DB hits: BLAST: pir:G63999;
transcription regulator GntR related protein ytrA -;
E=8e-17 pir:D72336; transcription regulator, gntR family -
Thermotoga; E=2e-14 pir:T36493; probable gntR-family
transcription regulator -; E=9e-11 COG: BS ytrA; COG1725
Predicted transcriptional regulators; E=8e-18 BS ytrA;
COG2188 Transcriptional regulators; E=8e-08 BH1164;
COG1725 Predicted transcriptional regulators; E=8e-08
PFAM: PF00325; Bacterial regulatory proteins, crip f;
E=0.49 PF00392; Bacterial regulatory proteins, gntR;
E=9.4e-16"
Alignment Scores:
Pred. No.: 2 06e+03 Length: 259950
Score: 132.00 Matches: 162
Percent Similarity: 32.72% Conservative: 88
Best Local Similarity: 21.20% Mismatches: 301
Query Match: 3.74% Indels: 216
DB: 1 Gaps: 32
US-10-691-383-2 (1-676) x BX294136 (1-259950)
QY 30 LysGlyTTPArgGlnLeuGluGlySerLeuGlyPheHisProSerGluThr 49
Db 59781 CGCGGGTGAACCAACGACATGCAGGC-----TTTGGCGGATGCTCTGAACCAAT 58831
QY 50 ProTyrIleGlySerLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAsp 69
Db 58832 CA-TCGCTCAATGTTGTATGGCGATGACGACATCGCGGAACGCGGATGATGTCAACA 58890
QY 70 GlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArg-----85
Db 58891 AGGTCACTTGTTCAGCCATCTACAGCAGATTTTCGTTGGAGGAAGATTGGCTGGATTGC 58950
QY 86 -----IleAlaThrAlaLeuAlaValLeuAlaAlaProCys-----98
Db 58951 TCAACGGTGTGCGACGTCATTCGCGCGTGTTCGGACACGCTGCTCGCGGATGCGCAAG 59010
QY 99 ---LeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThr 117
Db 59011 ACACITTTGCGTGCAGCGGATGTA-GCCGAAGAAATATCTCT-----59051
QY 118 GlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspProThrAla 137
Db 59052 -----TTGCAACG-----ACCAACGCGGACAAAC-----GATCCGACGCTG 59087
QY 138 ProAsnArgArgAsp-----142
Db 59088 ATCGTCAGCACCGCAGCGCAATACAGTTACGTCGGCGGTTGGTTCGTCGACTTTTGACGCC 59147
QY 143 ---AsnValAlaPheAlaSerArgAspAlaAlaArgArg-----GluArgAsp 158
Db 59148 AGTGGTGTGTGATTCATCATGATGTTGGACGCTTCGTCAGCGGAGTCTTTGGCCACTGAC 59207
QY 159 GlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHis 178
Db 59208 GAAGCGGGAACATG-----GTTCGTCACGTGTCGGCAGACATTCGCGCGGATCGCT 59261
QY 179 LysSerLeuProHisAspGluLeu-----GlyGlnValThr 190
```

```
Db 59262 GATAGTACCAAGGCGACGGAAGTCAACAACTGGTTCAGCCGTTCCGGGTGTGTCTCACT 59321
Qy 191 AlaAspAspPheAlaileLeu-----GluAspCysileLeuAsnGlyAspPheSerile 208
Db 59322 GCCAAGACGAGAGATCTTTGGTGCACCGATCTCTCTCGAAGACGCGCGGACGAG 59381
Qy 209 CysGlu-----AspValProAlaGlyAspProAlaGly 219
Db 59382 GTTCCGAACACAAACGAATCTGGCAACCTGACCGCGGACGCAACCTTGTGGCAGCT 59441
Qy 220 ArgLeuValAsnProThrAlaAlaPheAlaileAspPheSerile 234
Db 59442 CGTCAGTTGATGCTACAGTGGCGGTTTCGATCAAAATGTGTGGCGCATTCGGCGCGCG 59501
Qy 235 ---ProAlaPheSerAlaThrIleProProValProThrLeuSerSerProGluLeu 253
Db 59502 ATTGGTGGATCGATGACCGGTCGAGAGCTTCGACCGCGCGGCAACCCACGACGCA 59561
Qy 254 AlaAlaGlnLeuAlaGluLeuTyTrpMetAlaLeuAlaArgAspValProPheMetGln 273
Db 59562 GGAAGGAAGCAGCGAGATCTCTCAACTGGATATCGAAACGCGTTTCGGGTTT---GAC 59618
Qy 274 TyGlyThrAspGluIleThrThr----- 282
Db 59619 AACGGTTTACCGTGATGACGTCATCCCGATGGAGCTGAAGCGATTGATCGAACACGGC 59678
Qy 283 ---AlaAlaAlaAsnLeuAlaGly-----MetGlyGly 292
Db 59679 GTTGGCGGACCGATACCGCGGCAACCGCACACCGAGTCAGTTCCCGCAAAATCGGTGT 59738
Qy 293 Phe-----ProAsnLeuAspAlaValSerIleGlySerAspGlyThrVal 307
Db 59739 GCCAAGTTTCAGCTTTGATCAAAACGAAACAGCAATGTGTGGACGCTTCGCGCAACGTC 59798
Qy 308 AspProPheSerGlnLeuPheArgAlaThrPheValGlyValGluThrGlyProPheVal 327
Db 59799 GTTCAGGAAGCCGACGATTCGAGCTTCGCATTCGTCGACGATAACGGTGAATCATC 59858
Qy 328 SerGlnLeuLeuValAsnSerPheThrIleAspAlaileThrValGluProLysGlnGlu 347
Db 59859 GACGCTTGTGGCAATGCCAATTCGCGC-----ATCCCAATCCGGA 59906
Qy 348 ThrPheAlaProAspLeuAsnTyTrpMetValAspPheAspGluTrpLeuAsnIleGlnAsn 367
Db 59907 ATTGGTATCGTGACTTTGGACTTCCTCGCT-----GGGATCTTCTCG 59948
Qy 368 GlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAla 387
Db 59949 GCGCGTTCCACCGATTGGAGGC----- 59969
Qy 388 ArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyArgGlySer 407
Db 59970 -----GATGGCTACCGGTTCCCGGCTTCGGTGATCC 60002
Qy 408 LeuIleLeuLeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAsp 427
Db 60003 ATGACTCATCTTTCGTCGCTGTTGTTCG---GATGGTCGACGGATGTCTTCACGACC 60059
Qy 428 SerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyTrpPheArgLeuIleG 447
Db 60060 GGTAAAGGACGAGGATGCTGGTGGCTGAATACCTGTCGCCAACCAATCCGGATGGTGGG 60119
Qy 447 Y-----AlaAlaGluLeu-AlaGlnArgAlaSerCysTyTrpGlnLysTrpG 462
Db 60120 ACCGCTCTTATCGAGAGCGGACACTTCGGCAGTTCAAGTATCAAGTATCCAAATTTG 60179
Qy 462 InValHisArgPheAlaArgProGluAla---LeuGlyGlyThrLeuHisAsn-ThrIle 480
Db 60180 GCGGTACCAATGACAGATCTCTGACGCAATCTCGATCGACAGTTCCTCAAAATCTCCGTT 60239
Qy 481 AlaGlyAspLeuAspAlaAsp---PheAspIleSerLeuLeuGlu-----AsnAsp 496
Db 60240 GCCGGCACCTTCGAAACGGCGGTGTTCGAGAAATCGCGCGGGAATCGTGGCTCATGAC 60299
```

```
Qy 497 GluLeuLeuLysArgValAlaGluIleAsnAlaGlnAsnProAsnAsnGluValThr 516
Db 60300 CCGCAAAACCCAGGAGTCTTCTTCCAGAACTCCGATCCCAACACCATTTGTT--- 60350
Qy 517 TyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisPro----- 531
Db 60351 -----GTGCTGGACATTGGCCGATCCCAACCTCACCAAGGAACCTTTCCCCC 60395
Qy 532 ---SerTyrProSerGlyHis-----AlaThrGlnAsnGlyAlaPhe 544
Db 60396 ATCAGCTTCCAGCGGACCGGTGGTGTGAACAGCGTTCGGGTTTCTGGTGGCATCGTT 60455
Qy 545 AlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnPro 564
Db 60456 GCAGTCGCGGTGTGCTGCCAAGCGCACATCATCTCTGCTGGAGTCTCTGTTT--- 60506
Qy 565 ValPheProSerAspAspGlyLeu-GluLeuIleAsnPheGluGlyAlaCysLeuThrTy 584
Db 60507 ---TTCAGCTCGACCGCAACTTATCTCGGAGTGTTCACCGTGGGCTCTGCCAGACTCA 60563
Qy 584 rGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyI 604
Db 60564 CTCA----- 60567
Qy 604 eHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGlu----- 618
Db 60568 -CATTTTCTCCC-GACGGGATGAAAGTGTCTGTCTGTGTGAAGCGCAACCGGACGATTT 60625
Qy 619 -----ThrIleThrVal-----Ar 623
Db 60626 GGAAGATCCCAATCCGACCATCGATCCCTTGGGTACCATCAGCTGATGATTCGACTCG 60685
Qy 623 gThrLeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPh 643
Db 60686 ATTGCGTCAAGACGGCTTCATCATCTCTCGAGCTGACCACTCGACTTCACCGCGTT 60745
Qy 643 eThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCy 663
Db 60746 TGATGGACAAGAAGATCGCTCGCTCAAGGTGTACGATCTTCCCGGCGCGCTCGGC 60805
Qy 663 sSer 664
Db 60806 GTCA 60809

RESULT 15
AX598603
LOCUS AX598603
DEFINITION Sequence 11 from Patent WO02088176.
ACCESSION AX598603
VERSION AX598603.1 GI:28398756
KEYWORDS
SOURCE Streptomyces platensis
ORGANISM Streptomyces platensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycetaceae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Farnet, C.M., Zazopoulos, E., Staffa, A. and Vang, X.
TITLE Genes and proteins for the biosynthesis of polyketides
JOURNAL Patent: WO 02088176-A 11 07-NOV-2002;
Ecopia Biosciences Inc. (CA)
FEATURES
source 1. .9579
/organism="Streptomyces platensis"
/mol_type="unassigned DNA"
/db_xref="taxon:58346"
/notes="subsp. rosaceus"

ORIGIN
Alignment Scores:
Pred. No.: 27 Length: 9579
Score: 131.00 Matches: 152
Percent Similarity: 30.12%
```

Best Local Similarity: 19.56%				Mismatches: 254			
Query Match: 3.71%				Indels: 289			
DB:				Gaps: 38			
US-10-691-383-2 (1-676) x AX598603 (1-9579)							
QY	5	AlaAlaAspThrThrArgGly---SerProMetPro-AspThrGlyValLeuArgLeuLe 23		QY	269	lProPheMetGlnTyrGlyThrAspGluIleThrThrThralaAlaAlaAsnLeuAlaGl 289	
DB	6726	GCCGACATGAGCAGCGCGCCGCGACCTCTGACCGGAGCACTTCTACGCGCGCTCGG 6785		DB	7605	GCCGATGCCGCGCACCCGCGCGCGAGAC-----GACGCGGTGCCATCTGTCGCGCGC 7655	
QY	23	uThrSerGluGlnArgAlaLysGlyTrpArgGlnLeuGluGlyLysSerLeuGl 43		QY	289	yMetGlyGlyPheProAsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPr 309	
DB	6786	CACCGCGGCTCGACGCGCGCGCTGTCGCGCGCAGCTGCAAGGGGTCAACGAACTCGA 6845		DB	7656	TGCCGAGCGGTTCCTCCGCGCGGAC-----GACCTGGACAC 7691	
QY	43	yPheHis-----ProSer-----GluThrProTyrIleLy 53		QY	309	oPheSerGlnLeuPheArg-----AlaThrPheValGly-----Va 321	
DB	6846	CGCCACCGGCTGCTCGTACGGGTGCGCGAACCGGCCATGTGCCAGNACCGCACAGCA 6905		DB	7692	CTTCTGGCAGCAGTGTGCGCGCGGAGGACCTGATCGCQACTACCCCGCGCACCGCTT 7751	
QY	53	sTyrLeuGluGlySerGluThrTrpLysLysValLysLeuProThrAspGlyIleSer-- 72		QY	321	lGluThrGlyProPheValSerGlnLeuLeuValAsnSerPheThrIleAspAlaIleTh 341	
DB	6906	GCAGTGCAT---CTCCGCGCTGGCGCTCTGTCGGCTG---ACCAGGGTGTTCAGCA 6959		DB	7752	CGACGGGGCCCTTACCGGAGGTGTGTCGCGCGGCGACTTCCGAGATTTCGCGCGCG 7811	
QY	73	-----AlaSerLysIleLeuGlyLysIleMe 81		QY	341	rValGluProLysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPheAspGl 361	
DB	6960	CGCGTGGGCGCGGCGCGACGCGCGGTGCGGTGCGATCCGTGAGGCGGAGCAGTG 7019		DB	7812	GATCGAG---GGCGTGGACCGCTTCGACGCGGACTTCTTCCACTGTGCGGTGGAGGC 7868	
QY	81	tAlaArgValArgIleAlaThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPh 101		QY	361	u-----TyrLeuAsnIleGlnAs 367	
DB	7020	GGAGCGACCGG-----GCGATGCTGTGGCGG----- 7050		DB	7869	GGAGCTGATGGACCCGCGACGCGGCTGGCCCTGGAGACCGTGTGGCGCGCTGGAGAA 7928	
QY	101	eAspGluValThrAlaSerGlyValPheProGluGluHisLysHisThrGly----- 118		QY	367	nGlyGly-----ProProAlaGlyProGluGluLeu-----As 378	
DB	7051	-----ACGTCCAGCGCGTCTTC-----CATCGGCTTCTTCGCA 7085		DB	7929	CGCGGCTACGCGCGCGCGCTCCCGAGAACACCGGGGTCTACTTCGGCGTCTCCGG 7988	
QY	119	-----GluGlyArgHisLeuGlnThrCysThrAsnSerAsp-----AspAlaLeuAs 134		QY	378	pGluGluLeuArgPheIleArgAsnAlaArgAspLeuAla----- 391	
DB	7086	CGAGGACGGCGCGTGTGGCGCGGTGCGAGGACCGCGAGTTCAACGCGGCGACCTGGA 7145		DB	7989	CAGCGACTACCAACCACTGTCTCAACGCCAGTGGGTGGCGACCCGACCGGCTTCACCGCCAC 8048	
QY	134	pProThrAlaProAsnArgArgAspAsnValAlaPheAla----- 147		QY	392	-----ArgValSerPheValAspAsnIleAsn----- 400	
DB	7146	GCCGGCACTCCCGGTGAGCGCGGACGCGCACTCGTGGCACTGCGGAGCGGTGCGGTCC 7205		DB	8049	CGGCAACGCCCACTCGATGCTGGGCCAACCGGATCTCTACGTCTGGACGTGCACGGGCC 8108	
QY	148	-----SerArgArgAspAlaAlaArgArgGl 156		QY	401	-----ThrGl 402	
DB	7206	GGTCTGGAGACGCGGTTGGTACGGGGAGTGCGCAGCTGCGAGCGCGTGGCGCGGA 7265		DB	8109	GAGCGAACCGCTGCACACGCGCTGTCCAGCTCGCTCGCTGCGCTGCACCGCGCGCTCGA 8168	
QY	156	uArgAspGlyThrGlyThrValCysGlnIleThrAsnGly----- 169		QY	402	uAlaTyrArg-----GlySerLeuIleLeuLeuGl 412	
DB	7266	GGCGAGCGCTCCGTGACCGTGGCGCGGTGCGGACGCGCGCGCGCTGCTGCGGTC 7325		DB	8169	GCACATCCGCTCGGCGCGATGCGAGATGGCCATCGCGGGGGTGTCAACCTGTGCTGAG 8228	
QY	170	-----GluThrAspLeuAlaTh 175		QY	412	uLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAl 432	
DB	7326	GCTGCGCGAGACGCTGCGCGACCTGCTCAAGTTCGACCTGGCGGACATCGACCTCGACAC 7385		DB	8229	CGTGGACACCTTCGCGCGCAGCAGCATGGCGGGC---ATGCTCAGCCCCCGCGCGCTG 8285	
QY	175	rMetPheHisLys---SerLeuProHisAspGluLeuGlyGlnValThrAlaAspAspPh 194		QY	432	aGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIleGly---AlaAlaGluLe 451	
DB	7386	GCACTTCCACGCTACGGCTTCGAGTCCATCGCGCTGGCCAACTGGCCTCGGAACCTCAA 7445		DB	8286	CAAGACCTTCTCCGCGCGCGGACGCGCTAGCTCCGCTCCGAGGGCGTGCCTCGCGGTGT 8345	
QY	194	eAlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAl 214		QY	451	uAlaGlnArgAlaSerCysTyrGlnLys-----TrpGlnValHisArgPh 466	
DB	7446	CGGCGTCTCGGACGCGACCTCACCCCGCGCGTCTTCTTCGAGTGTCTCCGAC----- 7497		DB	8346	GCTCAAGCGCTGCGCCAGGCGCAGCGGACGCGGACGCGCATCTCTGGGGCGTCTGTCGGGG 8405	
QY	214	aGlyAspProAlaGlyArgLeuValAsnProThrAlaAlaPheAlaIleAspIleSerGl 234		QY	466	eAlaArg----- 468	
DB	7498	-----ATCCGAGCTCGCGGAGTACTGTCTCGACCGCTACCGCTACCG 7535		DB	8406	CAGCGCGGAGAACCAACGCGCGCGCGCGGTTCGCTGACCGCCCCCAACGCGAAGCGCA 8465	
QY	234	yProAlaPheSerAlaThrThr-----IleProValProThrLeuSe 249		QY	469	-----ProGluAlaLeuGlyGlyThr-- 475	
DB	7536	CCCCGAGCTGAGCTCCCCACGAGCGCGCGACGCCCGCGCGGTGCGCGCGCCACCGCGCC 7595		DB	8466	GGCGCGCTTATCAGGACGCCATGCGCGCATCGACCCCGGACAGCATCGGCTACGTCGA 8525	
QY	249	rSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspVa 269		QY	476	-LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAs 495	
DB	7596	GTCCCCA-----GT 7604		DB	8526	GGCGCAGCGCACCGGACCGCGC---CTGGGCGACCGGTGAGGTCAACGCTTCGACAG 8582	

```
Qy      515 lThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSe 535
Db      8611 -----GGCGGCGCGCGCACGCGCGCGCGCGTG 8639
Qy      535 r-----GlyHisAlaThrGlnAsnGlyAlaPheAlaTh 546
Db      8640 CGCGCTCGGCTCGGTGAAGACCAACATCGGCGCACGCGAGTCGCGCGCGCGCGCTTGCGCCGG 8699
Qy      546 rValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro----- 562
Db      8700 AGTGCTGAAGGTGCTGCTGCCATG---CGTCACCGCGAGCTGCCGCGCGCGCTTGCACTG 8756
Qy      563 -----AsnProValPheProSerAspGlyLeuGluLeuIle 575
Db      8757 CGACCGGCTCAACCGCACCTGCGCTCGACGCGCGGATTGAGTGTGTA 8805
```

Search completed: September 19, 2004, 07:31:11
Job time : 8839.83 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:16:07 ; Search time 125.3 Seconds
(without alignments)
1702.240 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTTGSGMPDGTGVL.....SIDGDMCSGLVYTGVDCA 676

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3528	100.0	676	10	O82433 fucus disti
2	1131.5	32.1	646	10	Q7X9V1 laminaria d
3	1126	31.9	682	10	Q7X9V0 laminaria d
4	626	17.7	598	10	Q8LLW7 corallina o
5	625	17.7	598	10	O81959 corallina p
6	614.5	17.4	597	10	O81960 corallina p
7	133.5	3.8	8173	16	Q7UDU8 rhodopirell
8	128	3.6	706	16	Q7UWV2 rhodopirell
9	127.5	3.6	650	17	Q9HNK0 halobacteri
10	125.5	3.6	4210	16	O89313 bradyrhizob
11	124.5	3.5	7716	16	Q7UWZ8 rhodopirell
12	123.5	3.5	959	16	Q87J64 vibrio para
13	123.5	3.5	3798	2	Q9L8C6 polyanium
14	123	3.5	1939	2	Q84HP6 amycolatops
15	122.5	3.5	469	16	Q9FCJ6 streptomyce
16	120.5	3.4	6310	16	Q88PP2 pseudomonas

ALIGNMENTS

RESULT 1

ID	O82433	PRELIMINARY;	PRT;	676 AA.
AC	O82433;			
DT	01-NOV-1998 (TrEMBLrel. 08, Created)			
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Vanadium bromoperoxidase.			
OS	Fucus distichus.			
OC	Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.			
OX	NCBI_TaxID=3012;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Vreeland V., Ng K., Epstein L.;			
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF053411; AAC35279.1; -			
DR	HSSP; P81701; 1019.			
DR	GO; GO:0004601; F:peroxidase activity; IEA.			
DR	InterPro; IPR008934; AcPase_VanPerase.			
DR	InterPro; IPR000326; PA_PTPase.			
KW	Peroxidase.			
SQ	SEQUENCE 676 AA; 73344 MW; AA8A46F539A14EC7 CRC64;			

Query Match	100.0%;	Score 3528;	DB 10;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 2.4e-258;		
Matches 676;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLCHAADTTGSGMPDGTGVLRLTSEORAKGWRRLQLEGEKSLGPHSPSETPIKYLEGSET 60		
Db	1	MLCHAADTTGSGMPDGTGVLRLTSEORAKGWRRLQLEGEKSLGPHSPSETPIKYLEGSET 60		
Qy	61	WKVKLPDGTGSGSKILGKIMARVRIATALAVLAAPCLAFDEVTAGVPEEHKHTGEG 120		
Db	61	WKVKLPDGTGSGSKILGKIMARVRIATALAVLAAPCLAFDEVTAGVPEEHKHTGEG 120		
Qy	121	RHLQCTNSDADLPTAPNRDNNVAFASRRDARRDGTGTCQITNGETDLATMFHKS 180		
Db	121	RHLQCTNSDADLPTAPNRDNNVAFASRRDARRDGTGTCQITNGETDLATMFHKS 180		


```
Qy 116 HTGERHLQCTNSDDALDPTAPNRRDNVAFASRRDAARRRDRGTGTVCQITNGETD--- 172
Db 65 QPLLSGNVCRVDSLDFDPV---RAKVTL--KRLAIKAEISVSGPTCHVNNNGENVP 120
Qy 173 -LATMEKHSKPHDSLGQVTDADDFAILEDICILNGDFSICEVDVAGDPAGR-----LVNPTA 226
Db 121 LFAQYHKTLPHDKFGQVDEDAYKKLJECFTSDINECEKVPSG--AGRRGGAKLTNPLG 178
Qy 227 AFAIDISGPAFSATTIPVPTLSSPELAQAELLYMALARDVPFMQYGTDEITTTAAAN 286
Db 179 GTAQVGTGADSDNVFITTPDSLSLSEAQAQAEVYMWALLARDIPFGFEAKNDVYRLAAEN 238
Qy 287 LAGMGFPNLDVANSIGSDGTVPDSQLFRATFTVGVETGPFVSQLVNSFTIDAITVEPKQ 346
Db 239 LQSLPAFKGLN-IPRSEGGKIDPVTDLFTTWPCTVTPVVSQPMLSDFLDSIKVTPKA 297
Qy 347 ETAPDLNVYVDFEWINIONGPPAGPEELDELRNARDLARVSFVDNINTEAYRG 406
Db 298 DPLTPGVYDTAFQPDWLDVQNGASKLETTTDEENPRFIRNGRDATIALRDQLYTEAFRA 357
Qy 407 SLILLELGAFSRPGINGPFIDSRRQAGFVNFSGTSHYFRLIGAELAQRASCYQKWQVHRF 466
Db 358 ALILFTGALG--GEVGPYAEAEQOGFATFGEPHILTAMASASSSTRHAWYAKWQVHRM 415
Qy 467 ARPEALGTTLHNTAGDLADDFDISLENDLLEKRVAEINAAQNPNNEVTVLLPQAIQVG 526
Db 416 LRPEAYGALVHTLMRDVITPLPDSILRNTLLELRNVEVHRQNMNPDGKFTFLPMAAQQ 475
Qy 527 SPTHSPSPSHATONGAFATVVKALIGLDRCGECFPNPPVPSDDGLELNF-----G 579
Db 476 SPTHAPSPSHAINNGAYITALKAFGLGYEQKCFPNPVSNDGTGRIKIKPSGREIVG 535
Qy 580 AC-----LTYEGEINKLVANVAFGRQMLGIHYRFDGICQGLLIGETITVRLHQEL 629
Db 536 ECVNEKSKLVEGLTYEGELINKISANVLLGRSHIGVHRMDGVYALNGETSCVRLQOEL 595
Qy 630 MTFAE-----BATPFRLTGEVVKLFQDGTFSIDGDMCSGLVYTVGVADC 674
Db 596 PGLPEAREVEGKRRGIDIPATYKFRLYSGKILEYGRNLYKLDGKLCBEG-AFTGDDDFC 653
```

RESULT 4

```
Q8LLW7 ID Q8LLW7 PRELIMINARY; PRT; 598 AA.
AC Q8LLW7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase.
OS Corallina officinalis (Coral seaweed).
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=35170;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22117781; PubMed=12121762;
RA Carter J.N., Beatty K.E., Simpson M.T., Butler A.;
RT "Reactivity of recombinant and mutant vanadium bromoperoxidase from
the red alga Corallina officinalis.";
RL J. Inorg. Biochem. 91:59-69(2002).
DR EMBL; AF218810; ANA46061.1; -.
DR POB; IQH8; 05-JUL-00.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65458 MW; BC784E370D748F01 CRC64;
```

```
Query Match 17.7%; Score 626; DB 10; Length 598;
Best Local Similarity 31.4%; Pred. No. 1.4e-38;
Matches 176; Conservative 76; Mismatches 220; Indels 88; Gaps 15;

Qy 168 NGETDLATMFHKSUPHDELQ-VTADDFAILEDICILNGDFSICBDV---PAGDPAGRL-- 221
```

```
Db 50 NGDPSFTVSTKGLPHDDNGAIIIDPDFLAFVRAINSGDEKEIADLTGLGPARDPTGLPI 109
Qy 222 -----VNPTAAFAIDISGPAFSATTIPVPTLSSPELAQAELLYMAL 265
Db 110 WRSDLANSLVGRWENSAGLTFDLEGDAQSIAMPAPVLTSPELIAEIAELYLWAL 169
Qy 266 ARDVPF-----MOYGTDEITTTAAANLAGMGFPNLDVANSIGSDGTVPDSQLP 314
Db 170 GREIFSEFSPKNAEYIQPAIQNLGLEWFTNPMLGDPPEAIRRRRGVTV---GNLNF 226
Qy 315 RATFVGVTGTFPVSQLLV-----NSFTIDAITVEPKQET 348
Db 227 RGLPGSEVGPYLSQYIIIVGSKIGSATGNKTLVSPNAADDFGEIAYGSITISQVRRI 286
Qy 349 FAPDLNVYVDFEWINIONGPPAGPEELDELRNARDLARVSFVDNINTEAYRGSL 408
Db 287 ATGRDFMTDLKVLVDQDAADFGESEYEPGARLTIIRDLATVWHFDALYEAYLNACL 346
Qy 409 ILLELGAFSRPGINGPFIDSRR--QAGFVNFSGTSHYFRLI-GAELAQRASCYQKWQVH 464
Db 347 ILLANRVFPDPI--PFQEDKLDNDQVFNFGDAHVLSTVTEVATRALKAVRYQKFNIIH 404
Qy 465 RFARPEALGTTLH-NTIAGD---LDADFDISLENDLLEKRVAEINAAQN-----PNNE 514
Db 405 RRLRPEATGGLISVKNITAAEKSGSEVPFVDLAVELEDEILEKAEISNRKQNIADGDDPDD 464
Qy 515 VTLLPQAIQVGSPTHSPSPSHATONGAFATVVKALIG---LDRGECFPNPPVPSDD 570
Db 465 PSFLLPQAFAGSGFPFHSYSGSHAVNAGACVTILKAFDFNSFDIQDVF-----VDKDED 519
Qy 571 GLELINPEGACLTGYEINLVANVAFGRQMLGIHYRFDGICQGLLIGETITVRLHQELM 630
Db 520 KLVKSPFKGT-LTAVAGELNKLADNIAIGRNAGVHYFSDQFESILLGEQVAILGSEQL 578
Qy 631 TPAREATFPERLFTGEVVKL 650
Db 579 TYGENFFFNLPKPDGTTIQI 598
```

RESULT 5

```
O81959 ID O81959 PRELIMINARY; PRT; 598 AA.
AC O81959;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 1.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimomishi M., Kuwamoto S., Inoue H., Wiewer R., Ohehiro T., Izumi Y.,
Tanabe T.;
RT "Cloning and expression of the gene for a vanadium-dependent
bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87657; BAA31261.1; -.
DR HSSP; P81701; IQ19.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65313 MW; 88B81B678B7FACE0 CRC64;
```

```
Query Match 17.7%; Score 625; DB 10; Length 598;
Best Local Similarity 31.9%; Pred. No. 1.6e-38;
Matches 191; Conservative 75; Mismatches 233; Indels 100; Gaps 20;

Qy 141 RDNVAFASRRDAARRRDRGTGTVCQITNGE-----TD--LATMFHKSUPHDELG 187
```

```
Db 11 RAKASFDRVAAAEALN-ARGVPSFANGELLRYNPDNDTDFSTASTKGLPHDDNG 69
Qy 188 Q-VTADDFAILDCILNGDFSCEDV---PAGDPAGRL-----VNPT 225
Db 70 AIIDPDDFLAFVRAINSQDEKEIADLTGLPARDPETGLPIWRSDLANSELEVRGWENSS 129
Qy 226 AAFADISGPAFSATTIPPVPTLSSPELAQAELVYMLARVDVPF-----MOY 274
Db 130 AGLTDLGGDDAOSIAMPAPVLTSPELVAEIAELVYMLALGRETIEFSEFSPKNAEYIQF 189
Qy 275 GTDEITTTAAANLGMGGFFNLDAVSGSDGTVDPPFSQLEFRATPGVGVETGPFVSQLL--- 331
Db 190 AIDOLNGLEWENTPAKLGDPPEAEIRRRRGEVTV---GNLFRGILPGSEVGPYLSQYIIVG 246
Qy 332 ---VNSFTI-----DAITVEPKOETFPADLNTMVDDEWLNTQNG 368
Db 247 SKQIGSATVGNKTLVSPNADEFDGEIAYGSITISQVRVRIATPGDRDFTDLKVFLVQDA 306
Qy 369 GPPAGPELDEELRFIRNARDLARVSFVDNINTEAYRGSLLLELGAFSRPGINGPEIDS 428
Db 307 ADFRGFSYEPGARLIITIRDLATWVHFDALYEAYLNACLIILANGVPFDP--NLPPQQE 364
Qy 429 DR---QAGVFNFGTSHYFRLI-GAAELAQASCYQKWQVHRFARPEALGGLTLH-NTIA-- 481
Db 365 DKLDNDQVFNFGSAHVLSLVTEVATRALKAVRYQKFNHRRLEPEATGGLISVKNLAPQ 424
Qy 482 --GDLADFDLISLENDELKRVAEINAAQN-----PNNEVYLLPOAIQVGSPTSPHSYP 534
Db 425 KGESIFPEVLDVBEELGILEKAEISNRKQNIADGDDPDPFSLPMAFAEGSPFHPHSYG 484
Qy 535 SGHATQNGAFATVLKALIGLDGCGECPNPVPFSD---DGLNLNFGACITVEGETNKL 591
Db 485 SGHAAVAGACVTILKAF--FDSGIEI--DQVEFVDEKEDKLKVSFKGT-LTVAGELNKL 539
Qy 592 AVNVAFGRMLGHIHYRFDGIGLLGTTITVTLHQELMTFAEBATPEFRLTGEVVKL 650
Db 540 ADNIAIGRNAGVHYFSDQFESLLLGQVAIGILEQSLTNGENFFNLPKFDGTTIQI 598

RESULT 6
O81960 PRELIMINARY; PRT; 597 AA.
AC O81960;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 2.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimonishi M., Kuwamoto S., Inoue H., Wiewer R., Ohehiro T., Izumi Y.,
RA Tanabe T.;
RT "Cloning and expression of the gene for a vanadium-dependent
RT bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87658; BAA31262.1; -.
DR HSP; P81701; 1019.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase_Vanperase.
DR InterPro; IPR000326; PA_Prtase.
KW Peroxidase.
SQ SEQUENCE 597 AA; 65199 MW; EAD931497060D3EE CRC64;

Query Match 17.4%; Score 614.5; DB 10; Length 597;
Best Local Similarity 31.6%; Pred. No. 1e-37;
Matches 193; Conservative 70; Mismatches 223; Indels 125; Gaps 21;

Qy 141 RDNVAFASRRDAARRRDTGTVQCITNGE-----TDLATMFHKSPLHDELGQ- 188
```

```
Db 11 RAKASFDRVAAAEALN-ARGVPSFANGELLRYNCETGDPSPFIASFTKGLPHDDNGAI 69
Qy 189 VTADDFAILDCILNGDFSCEDV---PAGDPAGRL-----VNPTAA 227
Db 70 IDPDDFLAFVRAINSQDEKEIADLTGLPARDPETGLPIWRSDLANSELEVRGWENSSAG 129
Qy 228 AAFADISGPAFSATTIPPVPTLSSPELAQAELVYMLARVDVPFMOYGTDE---ITTTAA 284
Db 130 LTDLGEGDDAQSAMPAPVLPMSPELVAEIAELVYMLALGRETIEFSEFSPKNAAFIRSAI 189
Qy 285 ANLAGMGFFNLDA-----VSGSDGTVDPPFSQLEFRATPGVGVETGPFVSQL 329
Db 190 ERLNGLEWF-NTPAKLGDPPEAEIRRRRGEVTVGN-----LFRGILPGSEVGPYLSQ 239
Qy 330 LL-----VNSFTI-----DAITVEPKOETFPADLNTMVDDEWL 363
Db 240 YIIIVGSKIGSATVGNKTLVSPNADEFDGEIAYGSITISQVRVRIATPGDRDFTDLKVFL 299
Qy 364 NIQNGGPPAGPELDEELRFIRNARDLARVSFVDNINTEAYRGSLLLELGAFSRPGING 423
Db 300 DVQDGADFRGFSYEPGARLIITIRDLATWVHFDALYEAYLNACLIILANGVPFDP--NL 357
Qy 424 PFIDSDRQAG---FVNFGTSHYFRLI-GAAELAQASCYQKWQVHRFARPEALGGLTL--- 476
Db 358 PFQEDKLDNDQVFNFGSAHVLSLVTEVATRALKAVRYQKFNHRRLEPEATGGLISVN 417
Qy 477 -HNTIAGDLADFDLISLENDELKRVAEINAAQNNEV-----TYLLPQAI 523
Db 418 KKSFLAGS-----DIIFPEVSELVEELSSILDVAESNEKQNRADGIVSPDKSFLLPMAF 472
Qy 524 QVGSPTSPHSYPSGHATQNGAFATVLKALIG---LDRGCECFNPVPFSDGLELINFE 579
Db 473 AEGSPFHPSCGSHAVVAGACVTILKAFDANFQIDKVE-----VDTDEKLKVSFKG 527
Qy 580 ACITYEGEINKLVNVAFGRMLGHIHYRFDGIGLLGTTITVTLHQELMTFAEBATPE 639
Db 528 T-LTVAGELNKLADNVAIGRNAGVHYFSDQFESLLLGQVAIGILEQSLTNGENFFN 586
Qy 640 FRLFTGEVVKL 650
Db 587 LPRFEDGTTIQI 597

RESULT 7
Q7UDU8 PRELIMINARY; PRT; 8173 AA.
AC Q7UDU8;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Probable aggregation factor core protein MAFp3, isoform C.
GN RB11769.
OS Rhodopirella baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294153; CAD79309.1; -.
KW Complete proteome.
SQ SEQUENCE 8173 AA; 826160 MW; B7DD4ADC483AC614 CRC64;

Query Match 3.8%; Score 133.5; DB 16; Length 8173;
Best Local Similarity 21.9%; Pred. No. 19;
```



```
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005097; AAC20220.1; -.
DR PIR; H84356; H84356.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR00873; AMP-bind.
DR Pfam; PF00501; AMP-binding.1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP_BINDING.1.
KW Ligase; Complete proteome.
SQ SEQUENCE 650 AA; 70748 MW; 2AF0B70543983853 CRC64;

Query Match 3.6%; Score 127.5; DB 17; Length 650;
Best Local Similarity 22.6%; Pred. No. 0.92; Indels 109; Gaps 22;
Matches 97; Conservative 48; Mismatches 176;

QY 10 RGSMPMDTGVLRLITSEORAKGWR-----RRQLEGEKSLGFHP-SETPYIKYLEGSETWK 62
DB 237 RFGPRDRGAPAVTSDSRHLSPLPLAHLVLERLSGH-----FLPLASGAQVCYAESPTLR 292
QY 63 K---VKLPTDGI SASKILGKIMARVRIATATAVLAAAPCLAFDEVTVASGVFPPEKHGTGE 119
DB 293 EDFGLPEPTSATSPRVYEKLYDTIREQAA-----DSPVSEVRF-EWATRVGR 339
QY 120 GRHLQTCNSDDALDPTAPNR-RDNVAFASRRDAARRD-----GTCTVQITNGETDLA 174
DB 340 AAH-----TTDDPGIGLRAAALADTLVFSIDRALGGEIFFISGGGSL-----SPELC 389
QY 175 TMFH-KSLPHDELQGVTTADDFAILEDCLNGDFISCEDVPAGDPAGRLVNPNTAAFAIDIS 233
DB 390 ALFHGMGLP-----LLEG-YGLTETSPV-----LAVNPYEDPIVGTI 425
QY 234 GPAPSAITIPVPTLSSPE-----LAAQLAELYWALARDVPFMQYGTDEI 279
DB 426 GPPVTDTELTVDETIAISPEQRQRCGGAAGELLARGPQVFDGYW-----GLPDA 473
QY 280 TTTAAANLAGMGFPNLDVSGISDGTV---DPPSOLFR-ATFVGVETGPFVSQLLVNSF 335
DB 474 TDAFVTRGKEWFRGTGDVVELRPGYVRFLEKAKQLLTSTGKNVAPGPIEDAFASPL 533
QY 336 TIDAITVEPKQETFAPDLNTVMVDFD---EMLNITQNGGPPAGPEELOBELAFIRNARDLAR 392
DB 534 VAQAMVVGDCQKFSAIL--VPNFDVSEWAASQEIALP-----DDRDAICDERVRAR 585
QY 393 V-SFVDNINT 401
DB 586 IQSAVDDVNT 595

RESULT 10
ID Q89UJ13 PRELIMINARY; PRT; 4210 AA.
AC Q89UJ13;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE B115471 protein.
GN B115471.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Igesawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
```

```
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005955; BAC50736.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003644; Calx beta.
DR InterPro; IPR000169; SHPOT acsite.
DR Pfam; PF03160; Calx-beta.1.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
KW Complete proteome.
SQ SEQUENCE 4210 AA; 428715 MW; D6D3B6879B95BAAF CRC64;

Query Match 3.6%; Score 125.5; DB 16; Length 4210;
Best Local Similarity 20.6%; Pred. No. 27;
Matches 166; Conservative 93; Mismatches 273; Indels 275; Gaps 45;

QY 6 ADTTRGSPMPD-TGVRLRLITSEORAKGRRQLRGEKSLGHPSE--TPYIKYLEGSETWK 62
DB 1080 ADETQGNQANDVTGPIQIFA-----GVQNPQGVDPDEPGTP-LAYATSSGAAL 1125
QY 63 KVKLPTDGISASKILGKIMARVRI-----ATALAVVLAAPCLAFDEVTVASGVFP 111
DB 1126 SI-LPYFGVDPGPPAAAITVSLKSSNFGVTLTDGTAISLSDSAGRVGTGTDNVHP 1184
QY 112 EEKHHTGEGRHLCCTNSDDALDP-----TAPNRDNVAFASRRDAARR 155
DB 1185 E-----LSGTTAFAVAIDPLSGKVSVEYLSLHSDSNPNQVSLAAGSISA-- 1232
QY 156 ERDGTGTVCQITNGETDLATM-----FKHSLPHDELQGVTTADDFAILEDCLNGD 205
DB 1233 -----VVTIKDGDGTASADISSKIHFNDDGDPALTIGTGTTPPELVAL---ALNLD 1282
QY 206 FSICED-----VPAGDPAGRLVNPNTAAFAIDISGPAPSAITIPVPTLSSPEL---A 254
DB 1283 ETIGHDGSPPDQVNLDEGESGGVNGTA---DDTGNATVTTTDTATSTAQIGELMTGA 1338
QY 255 AQLAELYWALARDVPFMQYGTD-----EITTTAAANLAG---MGGF-PNLDAVSTGSD 304
DB 1339 GLLAFLPTNAV-----IDYGADGAPALVNPVTSALSIVLTGAPEAGGIKTNLTAL-AD 1391
QY 305 GTVDPFSQLFRATFVGVETGPFVSQ-----LLVNSFTIDAITVEPKQETFA 350
DB 1392 SLLEGTSEQHRTVWLTIEGQIVGRVGHDTLSTADDVWLRSLNSTDPATAQLVVDQFL 1451
QY 351 PDLNVMVDEWLNQNGGPPAG-----PEELDEELRFR----- 385
DB 1452 P-----IDHD-----ASEPAGAQLPENPSLFDSEISLTLTATAGOSVGKLVTVTVDG 1499
QY 386 --NARDLARVSVDNINT-----EAYRG-----SLILLELGAFSRPGINGPFI 426
DB 1500 GDHVTKSATATVIDHSNSIVSIDDGPTTISASQDCVDTPLSYVGKAGNVDERGLDGP-S 1558
QY 427 DSDRQAGFVNGFTSHYFRIG--AELAQRASCYQKQVHRFARPBALGGTILHNTIAGD- 483
DB 1559 DQDVLMSGVGQGSNPANITNTGSDIGVGNQSGIEGHEVH-----GG---NDLAAEI 1607
QY 484 LDADF--DISLLENDELLKXVAEINAAQ-----NPNNEVTVLLPQALQVGSPTHPY 533
DB 1608 LRLDFVNNVTFPANTFAYNGHYEVDAASTFTHIOGNPNSNTATVF---VQV----- 1655
QY 534 PSGHATONGAFATVLKALIGDRGGCEFPNPV-----FPSDDGLELINF- 577
DB 1656 --FNANDNDFT-----DDGN---PLPITLNDVTVTGDASYTKIPVYDGNLTGVV 1701
QY 578 -----EGACLTVE--GEINKLAV-----NVAFG-----RQMLGIHYRFR--D 609
DB 1702 IVGLNEGATVTVDTSDNDFNLVISNDVAFSTPNGAPLTPVDPNDSQELNVDGHPFSVS 1761
QY 610 GIQGLLGETITVRLHQLMTFAEEA 636
DB 1762 GISSVVVCAFA-TLSVTHDESAGFAPOS 1787
```



```
QY 490 ISLENDLL-KRVAEINAAQNPNNEVYLLPQAIQVGSPTSPSYSP-----SGH 537
Db 757 VNLNRGELQXNVIAQSDDTPTGFCYLAESINDKGRSNPSQPCVCGEIRIKLDGE 816
QY 538 ---ATQNGAFATVLKALIGDRGCECPNPV 565
Db 817 YQPLKDNVSVTVSKD-SGIVKGAESFTAPV 846

RESULT 13
Q9L8C6
ID Q9L8C6 PRELIMINARY; PRT; 3798 AA.
AC Q9L8C6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Polyketide synthase.
GN EP0D.
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=So ce90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schupp T., Ono M., Zirkle R.E., Milnanow M., Cyr D.D.,
RA Nowak-Thompson B., Engel J.M., Toupke C., Stratmann A., Goff J.M.;
RA Gorlach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT epothilones A and B from Sorangium cellulosum So ce90.";
RL Chem. Biol. 7:97-109(2000).
DR EMBL; AF210843; AA226922.1; -.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR InterPro; IPR001601; Methyltransf.
DR InterPro; IPR002055; NAD BS.
DR InterPro; IPR006163; Pp Bind.
DR InterPro; IPR000051; SAM bind.
DR Pfam; PF00698; Acyl transf; 2.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt_C; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS00075; ACP DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 2.
KW Phosphopantetheine, Transferase.
SQ SEQUENCE 3798 AA; 405273 MW; 2033E57B254CC077 CRC64;

Query Match 3.5%; Score 123.5; DB 2; Length 3798;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

QY 158 DGTGTCQITNGETDLMTHFKSLPHDELQVTDADFAILEDG-----200
Db 2760 DGEHFVSQPLPEPLDAVL-----EAGRVFADLPVLFEWKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDFISICDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILFPGGSFDMARIYRDSPIARYSGIVRGVESAARVAVPSCMSILEIGTGTATT 2873
QY 242 IPPVPTLSSPELAQLAEYWMALA-----RDVPFMQYGTDEITTTTAAANLAGMGGFNP 295
Db 2874 AAVLPVLLPQRTVEHFTDVSPLFLARAEQRPRDYPFLKYGLDQDQEPAGQVHAQRFDV 2933
QY 296 LDAVSGSDGTVDPSQLFRATFVGVTGTFVSVOLLVNSFTIDAITVEPKQETPAP-DIN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNIQGGPPAGPELDELRF-----IRNARD-----LARVSFVDNI-----399

Query Match 3.5%; Score 123.5; DB 2; Length 3798;
Best Local Similarity 21.9%; Pred. No. 32;
Matches 109; Conservative 51; Mismatches 185; Indels 153; Gaps 22;

QY 158 DGTGTCQITNGETDLMTHFKSLPHDELQVTDADFAILEDG-----200
Db 2760 DGEHFVSQPLPEPLDAVL-----EAGRVFADLPVLFEWKFAGERLADVLTKTLA 2813
QY 201 --IL--NGDFISICDVPAGDP-----AGRLVNPTAAFAIDISGPAFSATT 241
Db 2814 LEILFPGGSFDMARIYRDSPIARYSGIVRGVESAARVAVPSCMSILEIGTGTATT 2873
QY 242 IPPVPTLSSPELAQLAEYWMALA-----RDVPFMQYGTDEITTTTAAANLAGMGGFNP 295
Db 2874 AAVLPVLLPQRTVEHFTDVSPLFLARAEQRPRDYPFLKYGLDQDQEPAGQVHAQRFDV 2933
QY 296 LDAVSGSDGTVDPSQLFRATFVGVTGTFVSVOLLVNSFTIDAITVEPKQETPAP-DIN 354
Db 2934 IVAANV-IHATRD-----IRAT-----AKRL-----SLLAPGGLL 2963
QY 355 YMVD---FDEWLNIQGGPPAGPELDELRF-----IRNARD-----LARVSFVDNI-----399

RESULT 14
Q84HP6
ID Q84HP6 PRELIMINARY; PRT; 1939 AA.
AC Q84HP6;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PKSE.
GN PKSE.
OS Amycolatopsis orientalis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
OX NCBI_TaxID=31958;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 43491;
RX MEDLINE=22447897; PubMed=12536216;
RA Zazopoulos E., Huang K., Staffa A., Liu W., Bachmann B.O., Nonaka K.,
RA Alelet J., Thorson J.S., Shen B., Farnet C.M.;
RT "A genomics-guided approach for discovering and expressing cryptic
RT metabolic pathways.";
RL Nat. Biotechnol. 21:187-190(2003).
DR EMBL; AF546139; AA025836.1; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR000794; Ketoacyl synth.
DR Pfam; PF00698; Acyl transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
SQ SEQUENCE 1939 AA; 203659 MW; F0BB981562210E09 CRC64;

Query Match 3.5%; Score 123; DB 2; Length 1939;
Best Local Similarity 22.6%; Pred. No. 12;
Matches 146; Conservative 59; Mismatches 249; Indels 192; Gaps 32;

QY 4 HAADTTTGSPMPDGTGVLRLTSTQRAK---GWRQLEGEKSLGFHPSETPYIKYLESET 60
Db 606 HADDIYRAANL-STGADQVATDVAQPRIVTGSAGLRVLKSLGTEATV--TGHSLGELT 662
QY 61 ---WKKVKLPDGISASKILGKIMARV-----RIATALA-----91
Db 663 ALHWGGAALTEREVLUKIAKIRGKYNATASDGDGMAAIAATPSPAEGLAEGEEVVIAGNA 722
QY 92 ---VLAAPCLAFDEVTA-----SGVFPFEEHKHTGEGRHQ---124
Db 723 PEQTVLSGPAEADRVVARARABEGVTAARINVSHTAFSPAVVPAEAEMTGLAIDAIDFARL 782
QY 125 ----TCTNSDDALDPTAPNR---RDNVAFASR-RDAARRRDGTGTVCQITNGE-----170
Db 783 DRPVPVSTVTGDLVHAEDLRLDLRDQVLPVRFREAAKVAERSDLVIEVGPGRVLTGLL 842
```

Search completed: September 19, 2004, 03:58:47
Job time : 129.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:14:52 ; Search time 7.02288 Seconds
(without alignments)
1468.042 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFLIGABLAQR.....GLLGLTITVTLHQLBLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	935	90.0	557	1 PRXV_ASCNO	P81701 ascopyllum
2	85.5	8.2	609	1 PRXC_CURIN	P49053 curvularia
3	85.5	8.2	1819	1 GCP6_HUMAN	Q96rt7 homo sapien
4	82.5	7.9	330	1 Y374_METJA	Q57819 methanococ
5	82	7.9	567	1 RST2_SCHPO	P78871 schizosacch
6	79.5	7.7	198	1 YBJG_ECOLI	P75806 escherichia
7	79.5	7.7	419	1 MUA2_STRP8	Q8p097 streptococ
8	79.5	7.7	419	1 MUA2_STRPY	Q99z78 streptococ
9	79.5	7.7	922	1 YB29_HUMAN	Q9hb77 homo sapien
10	78.5	7.6	205	1 Y236_THEAC	Q9hlj2 thermoplas
11	77.5	7.5	454	1 SR54_AQUAE	O67615 aquifex aeo
12	77.5	7.5	940	1 PTGA_MYCPN	P75569 mycoplasma
13	77	7.4	495	1 TOLC_ECOLI	P02930 escherichia
14	76	7.3	394	1 FKG_BACME	P24289 bacillus me
15	75.5	7.3	426	1 APEB_MYCLE	Q50022 mycobacteri
16	75.5	7.3	493	1 GALT_LACIC	Q98683 lactococcus
17	75	7.2	337	1 GLK_XYLFT	O87eg6 xyliella fas
18	75	7.2	774	1 STF_LAMED	P03764 bacterioph
19	74.5	7.2	260	1 OPRI_NEIMC	P10170 neisseria m
20	74.5	7.2	835	1 VIRL_AGRU1	P07167 agrobacteri
21	74	7.1	237	1 YEIU_ECOLI	P76445 escherichia
22	74	7.1	611	1 GLMS_RALSO	Q8y303 r glucosami
23	73.5	7.1	154	1 FMM_MORNO	Q9829 moraxella n
24	73.5	7.1	472	1 GCSB_THEVO	Q97c04 thermoplas
25	73.5	7.1	803	1 DCML_HYDPS	P19913 hydrogenoph
26	73	7.0	603	1 LEPA_SINY3	P74751 synechocyst
27	73	7.0	1163	1 Y222_HUMAN	Q92618 homo sapien
28	72.5	7.0	216	1 Y007_TREPA	Q83052 treponema p
29	72.5	7.0	304	1 STKG_MOUSE	Q88697 mus musculu
30	72.5	7.0	316	1 ARCC_BACLI	O86134 bacillus li
31	72.5	7.0	398	1 A23D_DROME	Q24093 drosophila
32	72.5	7.0	417	1 SAHH_METTH	O27673 methanobact
33	72.5	7.0	630	1 GIDA_PSEPU	P25756 pseudomonas

34	72.5	7.0	631	1 GIDA_PSEPK	Q88rw8 pseudomonas
35	72	6.9	337	1 GLK_XYLFA	Q9peg4 xyliella fas
36	72	6.9	488	1 GCSB_LISIN	Q92c04 listeria in
37	72	6.9	527	1 PPCK_BACHD	Q9k7g7 bacillus ha
38	72	6.9	857	1 GELA_DICDI	P13466 dictyosteli
39	71.5	6.9	466	1 STHA_YERPE	Q8za97 yersinia pe
40	71.5	6.9	499	1 UDPG_YEAST	P32861 saccharomyc
41	71.5	6.9	555	1 LEUI_BRUME	Q8yi13 bruceella me
42	71.5	6.9	615	1 MUTA_MYCTU	P71773 mycobacteri
43	71.5	6.9	660	1 FLR2_HUMAN	O43155 homo sapien
44	71.5	6.9	878	1 ACON_RICCN	Q92990 rickettsia
45	71	6.8	222	1 TOXK_PICFA	P19972 p salt-medi

ALIGNMENTS

RESULT 1
PRXV_ASCNO STANDARD; PRT; 557 AA.
AC P81701;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vanadium haloperoxidase (EC 1.11.1.-) (V-BPO).
OS Ascophyllum nodosum (Knotted wrack) (Brown seaweed).
OC Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaeeae;
OC Ascophyllum.
OX NCBI_TaxID=52969;
RN [1]
RP SEQUENCE, X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND FUNCTION.
RX MEDLINE=20013071; PubMed=10543953;
RA Weyand M., Wecht H.-J., Kiess M., Liaud M.-F., Vilter H.,
RA Schomburg D.;
RT "X-ray structure determination of a vanadium-dependent
haloperoxidase from Ascophyllum nodosum at 2.0-A resolution.";
RL J. Mol. Biol. 293:595-611(1999).
RN [2]
RP SEQUENCE OF 320-556 FROM N.A., SEQUENCE OF 326-341; 383-426; 471-479
AND 481-556, AND FUNCTION.
RX MEDLINE=96081028; PubMed=8564812;
RA Vilter H.;
RT "Vanadium-dependent haloperoxidases.";
RL (In) Sigel H., Sigel A. (eds.);
RL Metal ions in biological system-vanadium and its role in life,
RL pp.31-325-362, Marcel Dekker, New York (1995).
CC -1- CATALYTIC ACTIVITY: Halide + H(2)O(2) + H(+) = HOHal + H(2)O.
CC -1- COFACTOR: Vanadium.
CC -1- SUBUNIT: Homodimer; disulfide-linked.
CC -1- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-
PEROXIDASES.
DR PDB; 1O19; 10-JUN-00.
DR InterPro; IPR008934; Acpase_VanPerase.
DR InterPro; IPR000326; PA_PPase.
KW Oxidoreductase; Peroxidase; Vanadium; 3D-structure;
KW Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT DISULFID 3 3 INTERCHAIN (WITH C-41).
FT DISULFID 41 41 INTERCHAIN (WITH C-3).
FT DISULFID 77 86
FT DISULFID 441 462
FT DISULFID 544 555
FT ACT_SITE 411 411
FT ACT_SITE 418 418
FT METAL 486 486
FT CONFLICT 321 321 S -> D (IN REF. 2).
FT CONFLICT 341 341 K -> N (IN REF. 2).
FT CONFLICT 403 404 AI -> VY (IN REF. 2).
FT CONFLICT 407 408 GS -> T (IN REF. 2).
FT CONFLICT 409 409 P -> S (IN REF. 2; AA SEQUENCE).
FT CONFLICT 441 444 CYPD -> AIR (IN REF. 2).
FT CONFLICT 470 470 N -> K (IN REF. 2).
FT HELIX 15 37

```
FT HELIX 50 52
FT STRAND 53 53
FT TURN 56 57
FT STRAND 60 60
FT TURN 62 63
FT STRAND 66 66
FT HELIX 68 80
FT HELIX 83 87
FT TURN 88 88
FT HELIX 102 98
FT STRAND 105 108
FT TURN 118 119
FT TURN 129 130
FT HELIX 132 146
FT TURN 147 149
FT HELIX 152 154
FT TURN 155 157
FT HELIX 159 170
FT TURN 172 173
FT HELIX 174 176
FT TURN 178 179
FT STRAND 180 180
FT TURN 184 185
FT TURN 190 195
FT STRAND 197 197
FT TURN 200 203
FT HELIX 210 213
FT STRAND 216 218
FT TURN 219 220
FT STRAND 221 223
FT STRAND 227 227
FT STRAND 230 230
FT STRAND 236 236
FT HELIX 240 247
FT TURN 248 249
FT STRAND 258 258
FT STRAND 263 263
FT HELIX 268 277
FT TURN 282 293
FT TURN 294 295
FT TURN 299 300
FT HELIX 304 308
FT STRAND 312 313
FT TURN 316 317
FT STRAND 318 318
FT HELIX 320 329
FT TURN 330 333
FT HELIX 334 342
FT TURN 343 345
FT STRAND 347 347
FT HELIX 350 361
FT TURN 362 363
FT HELIX 371 374
FT TURN 375 375
FT HELIX 377 390
FT TURN 392 393
FT STRAND 402 402
FT TURN 405 406
FT HELIX 417 433
FT HELIX 435 438
FT TURN 439 439
FT STRAND 441 441
FT STRAND 446 448
FT TURN 450 451
FT STRAND 455 457
FT STRAND 464 464
FT HELIX 465 483
FT TURN 484 484
FT STRAND 519 522
FT TURN 524 525

Query Match 90.0%; Score 935; DB 1; Length 557;
Best Local Similarity 89.4%; Pred. No. 6.9e-81;
Matches 177; Conservative 12; Mismatches 9; Indels 0; Gaps 0;

QY 1 VNFQSHYFRLIGAAELAQRASYQKQVHRFARPEALGCTLHNTIAGDIDADFDISLLE 60
DB 316 VNFQSHYFRLIGAAELAQRSSWYQKQVHRFARPEALGCTLHNTIKELNADFDSLLE 375
QY 61 NDELLKRVABINAAQNPNEVTYLLPQAIQVGSPTSPSHATQNGAFATVVKALIGL 120
DB 376 NDELLKRVABINAAQNPNEVTYLLPQAIQVGSPTSPSHATQNGAFATVVKALIGL 435
QY 121 DRGCECPNPVFPSPDDGLSLINPEGACLTVEGINKLVNVAFCROWLGHYRFDGIQGL 180
DB 436 DRGDCYDPVPVDDGLSLIDFRGSLTPEGINKLVNVAFCROWLGHYRFDGIQGL 495
QY 181 LLGETITVTRLHOELMTFF 198
DB 496 LLGETITVTRLHOELMTFF 513

RESULT 2
PRXC_CURIN STANDARD; PRT; 609 AA.
AC P49053;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vanadium chloroperoxidase (EC 1.11.1.10) (VCPO) (Vanadium chloride
peroxidase).
GN CPO.
OS Curvularia inaequalis.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=38902;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=95262722; PubMed=7744081;
RA Simons B.H., Barnett P., Vollenbroek E.G.M., Dekker H.L.,
RA Muijsers A.O., Messerschmidt A., Wever R.;
RT "Primary structure and characterization of the vanadium
RT chloroperoxidase from the fungus Curvularia inaequalis.";
RL Eur. J. Biochem. 229:566-574(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=96133943; PubMed=8552646;
RA Messerschmidt A., Wever R.;
RT "X-ray structure of a vanadium-containing enzyme: chloroperoxidase
RT from the fungus Curvularia inaequalis";
RL Proc. Natl. Acad. Sci. U.S.A. 93:392-396(1996).
CC -1- CATALYTIC ACTIVITY: 2 RH + 2 chloride + H(2)O(2) = 2 RCl + 2
CC H(2)O.
CC -1- COFACTOR: Vanadium.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: The N-terminus is blocked.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85369; CAA59686.1; -;
```

DR PIR; S69334; S69334.
DR PDB; 1VNC; 08-NOV-96.
DR PDB; 1IDQ; 11-APR-01.
DR PDB; 1IDQ; 11-APR-01.
DR PDB; 1VNE; 11-AUG-99.
DR PDB; 1VNF; 11-AUG-99.
DR PDB; 1VNG; 11-AUG-99.
DR PDB; 1VNH; 11-AUG-99.
DR PDB; 1VNI; 11-AUG-99.
DR PDB; 1VNS; 11-AUG-99.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR003320; Cl_Perox.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF02328; Peroxidase; 1.
KW Oxidoreductase; Peroxidase; Vanadium; Chloride; 3D-structure.
FT ACT SITE 404 404
FT METAL 496 496
FT CONFLICT 454 454
FT HELIX 16 20
FT HELIX 22 40
FT TURN 41 41
FT HELIX 47 68
FT STRAND 79 79
FT TURN 80 81
FT HELIX 85 87
FT STRAND 88 88
FT TURN 93 94
FT HELIX 98 114
FT HELIX 131 147
FT TURN 149 150
FT TURN 154 155
FT HELIX 157 173
FT TURN 176 179
FT TURN 182 183
FT TURN 190 191
FT TURN 192 192
FT TURN 197 198
FT STRAND 202 207
FT TURN 208 209
FT STRAND 210 218
FT TURN 224 225
FT HELIX 226 229
FT TURN 245 246
FT HELIX 248 250
FT TURN 251 252
FT HELIX 254 268
FT STRAND 269 269
FT TURN 271 272
FT TURN 274 275
FT TURN 280 288
FT TURN 289 289
FT STRAND 294 294
FT TURN 295 297
FT STRAND 298 298
FT HELIX 300 314
FT TURN 315 315
FT HELIX 325 356
FT STRAND 358 358
FT HELIX 361 366
FT TURN 367 367
FT TURN 371 373
FT TURN 376 377
FT STRAND 380 380
FT STRAND 383 383
FT TURN 386 387
FT STRAND 393 393
FT HELIX 403 419
FT STRAND 420 420
FT TURN 421 423
FT STRAND 424 424
FT TURN 428 429
FT STRAND 436 440
FT HELIX 442 444

FT STRAND 449 451
FT TURN 457 458
FT HELIX 461 463
FT STRAND 466 468
FT STRAND 473 476
FT HELIX 479 492
FT TURN 493 494
FT HELIX 498 501
FT STRAND 504 507
FT STRAND 508 508
FT STRAND 510 510
FT TURN 514 515
FT STRAND 517 517
FT STRAND 519 519
FT TURN 521 522
FT STRAND 525 525
FT STRAND 527 527
FT HELIX 530 532
FT STRAND 539 539
FT STRAND 548 548
FT HELIX 552 565
FT TURN 566 567
FT HELIX 573 575
SQ SEQUENCE 609 AA; 67530 MW; A7B710DDF937D3E9 CRC64;
Query Match 8.2%; Score 85.5; DB 1; Length 609;
Best Local Similarity 21.7%; Pred. No. 2.7;
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;
Qy 71 INAAQNPNNEVYLLPOAIQVSGSPHPSYSGHATQNGAFATVLK-----115
Db 381 LGAPATNTDIPF-----KPPFPAYPSGHATFGGAVFQMYRRYNGRVGTWKDDP 431
Qy 116 -----ALIGLDRGCECPNPVPSPDGLLELI-----NFEAC-LTYEIGINKL 157
Db 432 DNTAIDNMISEELNGVNRDLRQPDYDTAPIEDQGIIVTRIVRHSWELMFENAIISRI 491
Qy 158 AVNVAFGRQMLGHIYRFD 175
Db 492 -----FLGVHVRFD 500
RESULT 3
GCP6 HUMAN STANDARD; PRT; 1819 AA.
ID GCP6 HUMAN Q9UGX3; Q9UGX4;
AC Q96RT7; Q9BY91; Q9UGX3; Q9UGX4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
GN Gamma-tubulin complex component 6 (GCP-6).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RX MEDLINE=21551508; PubMed=11694571;
RA Murphy S.M., Preble A.M., Patel U.K., O'Connell K.L., Dias D.P.,
RA Moritz M., Agard D., Stults J.T., Stearns T.;
RT "GCP6 and GCP6: two new members of the human gamma-tubulin complex.";
RL Mol. Biol. Cell 12:3340-3352(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguely C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,

RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grahame D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leverhulme M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Stuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M.J.M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuizuma S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Hu Y., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren O., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J.J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisole S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Koef I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Edelmann M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Badam L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard K., Kedar D., Seroussi E., Franconi I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliahun Y., Wright H.,
 RA "The DNA sequence of human chromosome 22."
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE OF 33-1819 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=21156230; PubMed=11258795;
 RA Hirosewa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
 RT "Identification of novel transcribed sequences on human chromosome 22
 RL DNA Res. 8:1-9(2001).
 CC -1- FUNCTION: Gamma-tubulin complex is necessary for microtubule
 CC nucleation at the centrosome
 CC -1- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,
 CC GCP3, GCP4, GCP5 and GCP6.
 CC -1- SUBCELLULAR LOCATION: Centrosome.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q96RT7-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q96RT7-2; Sequence=VSP_001624;
 CC Note=No experimental confirmation available;
 CC -1- SIMILARITY: Belongs to the GCP family.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
 CC in positions 1371 and 1758.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF272887; AAK82968.1;
 CC EMBL; AL022328; CAB63046.1; ALT_SEQ.
 CC EMBL; AL022328; CAB63047.1; ALT_SEQ.
 CC EMBL; AB051456; BAB33339.1; ALT_FRAME.
 CC EMBL; HGNC:18127; TUBGCP6.
 CC GO; GO:0008274; C:gamma-tubulin ring complex; IDA.

DR GO; GO:0008017; F:Microtubule binding; IDA.
 DR GO; GO:0007020; P:Microtubule nucleation; IDA.
 DR InterPro: IPR007259; SPC97_Spc98.
 DR Pfam: PF04130; SPC97_Spc98; 1.
 KW Microtubule; Repeat; Alternative splicing.
 FT DOMAIN 1027 1269 9 X 27 AA TANDEM REPEATS.
 FT REPEAT 1027 1053 1.
 FT REPEAT 1054 1080 2.
 FT REPEAT 1081 1107 3.
 FT REPEAT 1108 1134 4.
 FT REPEAT 1135 1161 5.
 FT REPEAT 1162 1188 6.
 FT REPEAT 1189 1215 7.
 FT REPEAT 1216 1242 8.
 FT REPEAT 1243 1269 9.
 FT VARSPLIC 1724 1757 Missing (in isoform 2).
 FT /FTID=VSP_001624.
 FT CONFLICT 567 567 S -> L (IN REF. 3).
 FT CONFLICT 1377 1377 A -> T (IN REF. 2 AND 3).
 FT CONFLICT 1621 1621 L -> V (IN REF. 2).
 SQ SEQUENCE 1819 AA; 200455 MW; 923576544034594A CRC64;
 Query Match 8.2%; Score 85.5; DB 1; Length 1819;
 Best Local Similarity 28.9%; Pred. No. 11;
 Matches 43; Conservative 18; Mismatches 55; Indels 33; Gaps 9;
 QY 7 HYRLICAAELAQRASYQK--WQVHRFARPEALGGTLHTNTIAGDLDDADFDISLEND-- 62
 DB 763 HYSL--SAAARRE---QALMRIOR-----HLESLARL-----FLEDEKH 801
 QY 63 --ELLKRVAEINAAQNPNNVTVLLPQAIQVGSPTSPGSHATONGAFATVTKALIGL 120
 DB 802 IQEMLKRVSAHQPPDPVLLSVHPQVTPSPG-PEHPE--GGQGCDSGSAEQHSPAWDGW 858
 QY 121 DRGCEFPNVPFP---SDDGLLEINFEQA 146
 DB 859 NRPGLLTPOPLKPLAVGAGRGGLQQAEGA 887
 RESULT 4
 Y374_METJA STANDARD; PRT; 330 AA.
 AC Q57819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0374.
 GN MJ0374.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcales; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=9637799; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Klerlavage A.R., Dougherty B.A., Tomo J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii."
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: TO M.JANNASCHII MJ0201 AND TO H.INFLUENZAE AND
 CC H.SONNUS H10703.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

Db 124 DLHLKAFAMGVSEYEGENNNLSTN---GQKHGAHIYMDTVS---VGATI 169

RESULT 8

MUA2_STRPY STANDARD; PRT; 419 AA.

AC Q9Z78; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase 2 (EC 2.5.1.7)

DE (Enol)pyruvate transferase 2 (UDP-N-acetylglucosamine enolpyruvyl transferase 2) (EPT 2);

GN MUR2 OR MURZ OR SPY1358.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OX Streptococcus.

NCBI_TaxID=1314;

RP SEQUENCE FROM N.A.

RC STRAIN=SF370 / ATCC 700294 / Serotype M1;

RX MEDLINE=21192684; PubMed=11296296;

RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;

RA "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*."

RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).

CC -1- FUNCTION: Cell wall formation. Adds enolpyruvyl to UDP-N-acetylglucosamine (By similarity).

CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + UDP-N-acetyl-D-glucosamine = phosphate + UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine.

CC -1- PATHWAY: Peptidoglycan biosynthesis; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: Belongs to the EPSP synthase family. MurA subfamily.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL: AE006573; AAK34186.1; .

DR HSP: P33038; 1DLG.

DR HAMAP: MF_00111; 1.

DR InterPro: IPR005750; AcGlu_Tran_MurA.

DR InterPro: IPR001986; EPSP_synth.

DR Pfam: PF00275; EPSP synthase; 1.

DR ProDom: PD001867; EPSP syntase; 1.

DR TIGRFAMs: TIGR01072; murA; 1.

KW Transferase; Peptidoglycan synthesis; Cell wall; Cell division;

KW Complete proteome.

SQ SEQUENCE 419 AA; 44662 MW; 4BCC957A6EAP489C CRC64;

Query Match 7.7%; Score 79.5; DB 1; Length 419;

Best Local Similarity 23.8%; Pred. No. 6.2;

Matches 41; Conservative 31; Mismatches 59; Indels 41; Gaps 9;

Qy 37 ALGGTTLHTIA---GDLADFDISLEN-----DELLKRVAEINAAQNPNNVTVLLP 86

Db 17 AVSGAKNSVVALIPAILAD-DIVILDGVPASIDVDSLSIEIMELMGATVNYHGDTLEIDP 75

Qy 87 QATQVGSPTSPSGHATONGAPATVL-----KALIGLDRGEGCFPPVPSDD 136

Db 76 RGVO-----DIPMPYTKINSRASYFYFGLSLGRFGQAVGLPGGCDLGRPI----- 123

Qy 137 GLELINF--GACLTVEGINKLAVNVAFGROMLGIHYRFDGIGLLGETI 186

Db 124 DLHLKAFAMGVSEYEGENNNLSTN---GQKHGAHIYMDTVS---VGATI 169

RESULT 9

UB29_HUMAN STANDARD; PRT; 922 AA.

AC Q9HJ7; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Ubiquitin carboxyl-terminal hydrolase 29 (EC 3.1.2.15) (Ubiquitin thiolesterase 29) (Ubiquitin-specific processing protease 29)

DE (Deubiquitinating enzyme 29).

GN USP29.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=20414752; PubMed=10958632;

RA Kim J., Noshkov V.N., Lu X., Bergmann A., Ren X., Warth T., Richardson P., Koprana N., Stubbs L.;

RT "Discovery of a novel, paternally expressed ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4";

RL Genome Res. 10:1138-1147(2000).

CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.

CC -1- SIMILARITY: Belongs to peptidase family C19.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL: AF229438; AAG10401.1; .

DR Genew: HGNC:18563; USP29.

DR MEROPS: C19.040; .

DR GO: GO:0004221; Fubiquitin thiolesterase activity; TAS.

DR GO: GO:0006464; P:protein modification; NAS.

DR InterPro: IPR001394; Peptidase_C19.

DR Pfam: PF00443; UCH; 1.

DR PROSITE: PS00972; UCH_2_1; 1.

DR PROSITE: PS00973; UCH_2_2; 1.

DR PROSITE: PS50235; UCH_2_3; 1.

KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.

FT ACT_SITE 294 294 BY SIMILARITY.

FT ACT_SITE 831 831 BY SIMILARITY.

FT ACT_SITE 840 840 BY SIMILARITY.

SQ SEQUENCE 922 AA; 104156 MW; 122B56AE6FFF364 CRC64;

Query Match 7.7%; Score 79.5; DB 1; Length 922;

Best Local Similarity 24.5%; Pred. No. 17;

Matches 48; Conservative 12; Mismatches 59; Indels 77; Gaps 9;

Qy 24 YQKQVHRPAP-----EALGGTLHTIAGDLADFDI----- 56

Db 143 YQKPLFMSKSPHVKVKGILENGGKGQNTLSSDVQTNEDILKEDNPVPHKKYKTDLSKY 202

Qy 57 -----SL--LENDELLKRVAEI--NAAQNPNNVTVLLPQAIQ-----VGSPTSPSY 99

Db 203 IQSNRKNPSSLEDKDRDLKLGPSFNTNCGNPNLDETVLATQTLNAKNGLTSPLEPEH 262

Qy 100 PSGHATONGAFATVLKALIGLDR-----GGECFPNPVFPSS-----DGL- 138

Db 263 SQGDRPCN-----RAQVPLDSHQLOQGFNLTGNTCYMNAVLSLPAIPSPADLLT 315

Qy 139 -----ELINFEAGLCT 149

Db 316 QGVPEVIFPEALIMT 331

```

RESULT 10
Y236 THEAC
ID Y236_THEAC STANDARD; PRT; 206 AA.
AC Q9HJG2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein ta0236.
GN TA0236.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -1- SIMILARITY: Contains 1 AMMECR1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AL445063; CAC11381.1; -.
CC HAMAP; MF 00645; -.
CC InterPro; IPR002733; DUF51.
CC Pfam; PF01871; AMMECR1; 1.
CC ProDom; PD009671; DUF51; 1.
CC TIGRFAMs; TIGR00296; TIGR00296; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 206 AA; 22971 MW; 5E83B24049A40278 CRC64;
-----
Query Match 7.6%; Score 78.5; DB 1; Length 206;
Best Local Similarity 25.1%; Pred. No. 3.1;
Matches 45; Conservative 19; Mismatches 54; Indels 61; Gaps 10;
QY 52 ADFDISLLENDELKRVAEINAAQNPNVETLLPQAIQVSGSPHPSYPSGHATQNGAFA 111
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 AEVDISLDIGTKAV-RLRAAAASSYLRLNE---KLDP-----PPADPTFQEKH---GVFT 55
-----
QY 112 TV-----LKALIGLDRGCECPNPVPSDDG-----LELIN 142
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 56 TINTYPGNTLRGCGI-----PPEYYPILGEGIRSSIYAATEDPRFPMKIDEISHVT 108
-----
QY 143 FEGACLYEYGEINKLVN-----VAFGRQMLGIHYRFGDIOGLLIGETITVRLTQHE 194
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 FEVSILTQPEI-----TVPEDRPKAVHIGRD--GLIAYNGASGLLLPQVATEYRMNPE 162
-----
RESULT 11
SR54 AQUAE
ID SR54_AQUAE STANDARD; PRT; 454 AA.
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal recognition particle protein (Fifty-four homolog).
GN FFH OR AQ.1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Shead M.A., Keller M., Anjaj M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -1- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC -1- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -1- SIMILARITY: Belongs to the GTP-binding SRP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE000753; AAC07579.1; -.
CC PIR; E70448; E70448.
CC HSP; O07347; IFFH.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR000897; SRP54.
CC InterPro; IPR004125; SRP54_SPB.
CC InterPro; IPR004780; SRP_sub.
CC Pfam; PF00448; SRP54; 1.
CC Pfam; PF02881; SRP54_N; 1.
CC Pfam; PF02978; SRP_SPB; 1.
CC ProDom; PD000819; SRP54; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMs; TIGR00959; ffh; 1.
CC PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 289 G-DOMAIN.
FT DOMAIN 290 454 M-DOMAIN.
FT NP_BIND 102 109 GTP (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
SQ SEQUENCE 454 AA; 50911 MW; 9F13FA8C83A0558 CRC64;
-----
Query Match 7.5%; Score 77.5; DB 1; Length 454;
Best Local Similarity 23.9%; Pred. No. 11;
Matches 43; Conservative 26; Mismatches 40; Indels 71; Gaps 10;
QY 12 IGAAELAQRCASYQKQVHRFAPEALGGLTHTNTIAGDLDADFDISLLENDELKRVAEI 71
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 160 LGAVEIARRA-----VKR-AKEESVDYLLDT-AGRLHVD-----EELMKELQEI 202
-----
QY 72 NAAQNPNVETLL-----PQAIQVSGSPHPSYPSGHATQNGAFAFATVLLKALIGLDRGCEC 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 203 KEVTNP-SEILYVADAMQQTALKTFT-----ERLGLTGVVITKQDGDARGG-- 251
-----
QY 127 FPNPVPFDDGLELINFEGACLYEYGEINKLVNVAFGROMLGIHYRFGDIOGLLIGETI 186
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 252 -----LALSV---KEVLGPPIKFIGV-----GEKI 273
-----
RESULT 12
PTGA MYCPN
ID PTGA_MYCPN STANDARD; PRT; 940 AA.
AC P75569;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

PTS system, glucose-specific IIABC component (EIIABC-GLC) (Glucose-permease IIABC component) (Phosphotransferase enzyme II, ABC component) (EC 2.7.1.69) (EII-GLC/EIIB-GLC).

GN PTC OR MPN207 OR MP624.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RA MEDLINE=97105885; PubMed=948633;

RX Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C., Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."

RL Nucleic Acids Res. 24:4420-4449 (1996).

CC -!- FUNCTION: This is a component of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS), a major carbohydrate active transport system. The IICD domains contain the sugar binding site and the transmembrane channel; the IIA domain contains the primary phosphorylation site (the donor is phospho-HPr); IIA transfers its phosphoryl group to the IIB domain which finally transfers it to the sugar.

CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein histidine + sugar phosphate.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: Contains 1 PTS EIIA domain.

CC -!- SIMILARITY: Contains 1 PTS EIIB domain.

CC -!- SIMILARITY: Contains 1 PTS EIIC domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE000060; AAB96272.1; -.

DR PIR; S73950; S73950.

DR HSP; P05053; 11BA.

DR InterPro; IPR001996; Pfam EIIIB.

DR InterPro; IPR003352; Pfam EIIA.

DR InterPro; IPR001127; PIR EIIA.

DR Pfam; PF00358; PIR EIIA 1; 1.

DR Pfam; PF00367; PIR EIIIB; 1.

DR Pfam; PF02378; PIR EIIIC; 1.

DR ProDom; PD002243; PIR EIIA; 1.

DR ProDom; PD001476; PIR EIIIB; 1.

DR TIGRFAMs; TIGR00826; EIIIB_glc; 1.

DR TIGRFAMs; TIGR00830; PTBA; 1.

DR PROSITE; PS00371; PIR EIIA 1; 1.

DR Phosphotransferase system; Sugar transport; Transferase; Phosphorylation; Transmembrane; Complete proteome.

KW Phosphorylation; Transmembrane; Complete proteome.

FT DOMAIN 1 7290

FT DOMAIN 7291 7446

FT DOMAIN 7447 7628

FT DOMAIN 7629 7730

FT DOMAIN 7731 940

FT TRANSMEM 43 63

FT TRANSMEM 83 103

FT TRANSMEM 112 132

FT TRANSMEM 175 195

FT TRANSMEM 209 229

FT TRANSMEM 238 258

FT TRANSMEM 265 285

FT TRANSMEM 487 507

FT TRANSMEM 515 535

FT TRANSMEM 537 557

FT TRANSMEM 564 584

FT TRANSMEM 598 618

FT TRANSMEM 671 683

FT MOD_RES 271

FT MOD_RES 683

FT MOD_RES 847 847 PHOSPHORYLATION (BY SIMILARITY).

SQ SEQUENCE 940 AA; 101619 MW; 44B836307FDA36EF CRC64;

Query Match 7.5%; Score 77.5; DB 1; Length 940;

Best Local Similarity 26.1%; Pred. No. 27;

Matches 48; Conservative 22; Mismatches 55; Indels 59; Gaps 11;

QY 30 HRFARP---BALGGTTLHTIAGDLDADFDISLLENDELLKRVASINAAQNPNVYLLP 86

DB 272 HAFVAPLWYTSAGSGLQEIYNQVWIRPDEFLSDN-----YVARVIGWVDNNSMWIIP 326

QY 87 QAI--QVGSPTH-----PSYSGHATQNGAFATVLKALIGLDRGCECFNPVP 133

DB 327 GALTNGQSGSTGNTMSKOLNALSAYMSKST---AFLT-WKDLV----- 367

QY 134 SDGGLLEINFEAGACLYEVEIKLAVN-VAFGRQMLGIHYRFDGIQG--LILGE---TI 186

DB 368 --DG-----LTPKGNFDMKWAENGLDGSNKIWLGLNGSILGKLLSDGNVYTI 415

QY 187 TVRT 190

DB 416 TFKT 419

RESULT 13

TOLC ECOLI

ID TOLC ECOLI STANDARD; PRT; 495 AA.

AC P02930;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Outer membrane protein tolc precursor.

GN TOLC OR MTCB OR MUKA OR REFI OR B3035.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=91016844; PubMed=2216730;

RA Niki H., Imamura R., Ogura T., Hiraga S.;

RT "Nucleotide sequence of the tolc gene of Escherichia coli.";

RL Nucleic Acids Res. 18:5547-5547 (1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=84015386; PubMed=6312426;

RA Hackett J., Reeves P.;

RT "Primary structure of the tolc gene that codes for an outer membrane protein of Escherichia coli K12.";

RL Nucleic Acids Res. 11:6487-6495 (1983).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474 (1997).

RN [4]

RP SEQUENCE OF 25-36.

RC STRAIN=K12 / EMG2;

RX MEDLINE=97443975; PubMed=9298646;

RA Link A.J., Robison K., Church G.M.;

RT "Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12.";

RL Electrophoresis 18:1259-1313 (1997).

RN [5]

RP SEQUENCE OF 25-29.

RC STRAIN=K12 / W3110;

RX MEDLINE=98291876; PubMed=9629924;

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC	EMBL; U15182; AAA62998.1; ALT INIT.
DR	EMBL; Z95151; CAB08404.1; ALT INIT.
DR	EMBL; AL583924; CAC31168.1; ALT INIT.
DR	EMBL; AL583924; CAC31168.1; ALT INIT.

FT	METAL	156	ZINC (POTENTIAL) .
FT	METAL	399	ZINC (POTENTIAL) .
SQ	SEQUENCE	426 AA; 45593 MW; 582ADDD29C3A0A53	CRC64;

Query local	7.5%; Score 73.5; DB 1; Length 420;
Best local Similarity	22.7%; Pred. No. 15;
Matches: 46; Conservative	26; Mismatches 68; Indels 63; Gaps 10;
Qy	11 LIGAEALQASCYQKWVHRFARPEALGGTL-----HNTIAGDLDADFDISL 58
	: : : : : : : : : : : : : : : : : : :
Db	222 LLSAPRLDNQVCY-----AGMEALLASVPHDCLPVLLPDPHEVGVSSTDRGARSNL 273
Qy	59 LENDELLKRVAEINAAQNPNVEYLLLPQAIQVS-----PTHPSPYSGHATQNGAEATVL 114
	: : : : : : : : : : : : : : : : : :
Db	274 LST--VLERI--VLAAGGGRDYLRLPSSLVSDMAHATHENYPECHPSH----- 322

Db 323 --LIEVNAQPVLKVHNRLRY-ATDGRSTAAAFVAC-----QQAGVRL 361

Qy 171 ---HYRFDGIQGLLLGETITVRT 190

Db 362 QRYEHRADRPCGSTIGFLASART 384

Search completed: September 19, 2004, 03:55:58
Job time : 9.0288 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 03:59:53 ; Search time 5444.35 Seconds
(without alignments)
3707.849 Million cell updates/sec

Title: US-10-691-383-2
Perfect score: 3528
Sequence: 1 MLCXAADTRGSPMPDTGVL.....SIDGDMCSGLVYTGVAQCQA 676

Scoring table:
BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10691383/runat_17092004_102656_1753/app_query.fasta_1.1230
-DB=EST -QPM=fastap -SUFFIX=rst -MINMATCH=0.1 -LOPCT=0 -LOPEXT=0
-UNITS=bites -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10691383 @CGN 1.1 4591 @runat_17092004_102656_1753 -ICPU=3
-NO MMAP -LARGEJOEY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	217	6.2	393	10	AW400475	AW400475 Lamdisest
2	199.5	5.7	503	9	AUI87248	AUI87248 AUI87248
3	199.5	5.7	538	9	AUI91942	AUI91942 AUI91942
4	146	4.1	503	9	AUI94816	AUI94816 AUI94816
5	138.5	3.9	441	9	AUI87606	AUI87606 AUI87606
6	117	3.3	1201	9	AL539002	AL539002 AL539002
7	115	3.3	2089	29	AY411102	AY411102 Homo sapi
8	115	3.3	2673	11	BC040379	BC040379 Homo sapi
9	114.5	3.2	889	12	BI836552	BI836552 603089514
c 10	113	3.2	369	14	CD063088	CD063088 MA1-0029U
c 11	113	3.2	936	11	CNS080UB	BX007359 Single re
12	110.5	3.1	701	14	CB552155	CB552155 MMSPP0080
13	110.5	3.1	2823	11	AK039670	AK039670 Mus muscu
c 14	110	3.1	1052	29	CNS03HC2	AL244091 Tetraacu
c 15	109	3.1	712	14	CB443557	CB443557 694466 MA
c 16	108	3.1	392	9	AUI89808	AUI89808 AUI89808
17	108	3.1	1168	12	BM456445	BM456445 AGENCOURT
18	106.5	3.0	668	28	BH86578	BH86578 LB00877A
19	105.5	3.0	3208	11	AK039695	AK039695 Mus muscu
20	105	3.0	3250	11	BC047878	BC047878 Homo sapi
21	104	2.9	3371	29	AY398790	AY398790 Homo sapi
c 22	103.5	2.9	901	29	CG285791	CG285791 OGWFL56TH
c 23	103.5	2.9	995	14	CF224754	CF224754 AGENCOURT
24	103.5	2.9	1680	12	BM925460	BM925460 AGENCOURT
c 25	103.5	2.9	3254	11	CNSLT1170	EX161390 human ful
c 26	103	2.9	3273	29	AY398791	AY398791 Pan trogl
c 27	102	2.9	1116	11	AK000257	CD500257 CDA44-A04
c 28	102	2.9	3117	14	AD045649	AD045649 Mus muscu
c 29	101.5	2.9	852	28	BZ709996	BZ709996 OGBAR23TC
c 30	101	2.9	1060	12	BM805002	BM805002 AGENCOURT
31	100.5	2.8	844	14	CD519980	CD519980 AGENCOURT
c 32	100.5	2.8	1323	28	BZ572188	BZ572188 msn2_2542
c 33	100	2.8	400	10	AW400432	AW400432 Lamdisest
c 34	100	2.8	787	28	BZ710005	BZ710005 OGBAR23TM
c 35	99.5	2.8	780	29	CG700975	CG700975 OGUFW58TH
c 36	99.5	2.8	1398	29	U90627	U90627 U90627 Azos
c 37	99	2.8	655	10	BE822718	BE822718 GM700018B
c 38	99	2.8	913	12	BG741433	BG741433 602632215
c 39	99	2.8	1054	12	BI414575	BI414575 602986154
c 40	98.5	2.8	584	12	BG897929	BG897929 HOA2-G-1
c 41	98.5	2.8	930	13	BQ939745	BQ939745 AGENCOURT
c 42	98.5	2.8	1027	29	CG687286	CG687286 ZMMBCC016
c 43	98.5	2.8	1125	10	BE307732	BE307732 601095594
c 44	98.5	2.8	1504	11	AY104439	AY104439 Zea mays
45	98	2.8	754	13	BU100257	BU100257 WHE3351_B

ALIGNMENTS

RESULT 1
AW400475
LOCUS
DEFINITION Lamdisest169est L.digitata sporophyte Lambda ZapII Laminaria
ACCESSTON AW400475
VERSION AW400475.1
KEYWORDS EST.
SOURCE Laminaria digitata
ORGANISM Laminaria digitata
Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
Laminariaceae; Laminaria.
REFERENCE 1 (bases 1 to 393)

AW400475 393 bp mRNA linear EST 07-FEB-2000
Lamdisest169est L.digitata sporophyte Lambda ZapII Laminaria
digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.


```

ACCESSION   AU191942
VERSION     AU191942.1
KEYWORDS    GI:31930088
SOURCE      Porphyra yezoensis
ORGANISM    Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1 (bases 1 to 538)
AUTHORS    Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
TITLE       COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
JOURNAL     J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES    Location/Qualifiers
             source
               1..538
               /organism="Porphyra yezoensis"
               /mol_type="mRNA"
               /strain="TU-1"
               /db_xref="taxon:2788"
               /clone="PFL023f08_r"
               /dev_stage="sporophytes"
               /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN
Alignment Scores:
Pred. No.:      2,84e-10      Length:      538
Score:          199.50      Matches:    57
Percent Similarity: 51.54%      Conservative: 10
Best Local Similarity: 43.85%      Mismatches: 54
Query Match:    5.65%      Indels:    9
DB:              Gaps:      4

US-10-691-383-2 (1-676) x AU191942 (1-538)
QY  429 AspArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIle---Gly 447
DB  91 GATCATCAGCAGGGTTTGCCACGTTTGGGGGCCCCCACCCTGTGACGCTGCACCGAG 150
QY  448 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlyStrpGlnValHisArgPheAla 467
DB  151 GTGCCACACGCGCACTCAAGCGGTCGCTTCCAAAGTTCAACGTCACCGCGGCTC 210
QY  468 ArgProGluAlaLeuGlyThrLeuHisAsn---ThrIleAlaGlyAspLeuAspAla 486
DB  211 CGGCCCGAGCGGTGGCGGCGCTCTCGACCGCTGCGCGCCGCGCCAGGACCGCTGGCG 270
QY  487 AspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuIysArgValAlaGluIleAsn 506
DB  271 CCAGTGGCCCGCTGGCAGAGGAGCGCGGCGAGACGCTCGCGCGGGGTGTGATGCCAAC 330
QY  507 AlaAla-----GlnAsnProAsnGluValThrTyr-----LeuLeu 519
DB  331 GCTGCCCGCCCTCGGGGGGCGAGTGGCGCGCGGTGCGCGCCGTTACGACAGTGTGCTGTG 390
QY  520 ProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThr 539
DB  391 CCAATGGCCCTTCCAGAGGGGCTCCCGCATGCACCCCTCTCTACGGCGCGGTCACGCCACC 450
QY  540 GlnAsnGlyAlaPheAlaThrValLeuIys 549
DB  451 GTTGGCGGCGCGTGCCTGACGGGTCTCAAG 480

RESULT 4
LOCUS      AU194816
DEFINITION Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
            cDNA clone PFL065g09_r 5', mRNA sequence.
ACCESSION   AU194816
VERSION     AU194816
KEYWORDS    GI:31921404
SOURCE      Porphyra yezoensis

```

```

ACCESSION   AU194816
VERSION     AU194816.1
KEYWORDS    GI:31935825
SOURCE      Porphyra yezoensis
ORGANISM    Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1 (bases 1 to 503)
AUTHORS    Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
            Tabata,S.
TITLE       COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
            OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
            FREQUENCY ANALYSIS
JOURNAL     J. Phycol. 39 (5), 923-930 (2003)
COMMENT     Contact: Erika Asamizu
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES    Location/Qualifiers
             source
               1..503
               /organism="Porphyra yezoensis"
               /mol_type="mRNA"
               /strain="TU-1"
               /db_xref="taxon:2788"
               /clone="PFL065g09_r"
               /dev_stage="sporophytes"
               /clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN
Alignment Scores:
Pred. No.:      0.000114      Length:      503
Score:          146.00      Matches:    42
Percent Similarity: 49.52%      Conservative: 10
Best Local Similarity: 40.00%      Mismatches: 43
Query Match:    4.14%      Indels:    10
DB:              Gaps:      3

US-10-691-383-2 (1-676) x AU194816 (1-503)
QY  177 PheHisIysSerLeuProHisAspGluLeuGlyGlnValThrAlaAspAsp-----Phe 194
DB  48 TTCAACAAAGGTCTGCAGCAACCTCTCGACGGGCTCTCTGGAGGACCCCGCGCCTTT 107
QY  195 AlaIleLeuGluAspCysIleLeuAsnGlyAspPheSerIleCysGluAspValProAla 214
DB  108 CGCGCCTTTGTCGCGGCATTGACAGCGGGACCCCGCGGAGCTTTGTCTGACGTGCCCAT 167
QY  215 GlyAsp-----ProAlaGlyArgLeuVal-----AsnProThrAla 226
DB  168 GCGGACCGCTTACCGCGCCCGCCACCGCGGGTGGCGTCTGGGAGTCTGGCGTGGCG 227
QY  227 AlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrIleProProValPro 246
DB  228 GGCGTGGCGTGTGACCTTACGGGCGCGGATGCGGAGGCGGTGACCATGCCCGCCCGCCG 287
QY  247 ThrLeuSerSerProGluLeuAlaAlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAla 266
DB  288 CGCGTGGGACGCGGNACTGGTGGCGGAGATGATTGAGGTCTACGCCATGGCGTCCGCG 347
QY  267 ArgAspValProPhe 271
DB  348 CGGGATGTGCGGTTT 362

RESULT 5
LOCUS      AU187606
DEFINITION Porphyra yezoensis TU-1 sporophytes Porphyra yezoensis
            cDNA clone PFL04hal_r 5', mRNA sequence.
ACCESSION   AU187606
VERSION     AU187606.1
KEYWORDS    GI:31921404
SOURCE      Porphyra yezoensis

```

ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
REFERENCE 1 (bases 1 to 441)
AUTHORS Asamizu,E., Nakajima,M., Kitade,Y., Saga,N., Nakamura,Y. and
Tabata,S.
TITLE COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS
OF PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG
FREQUENCY ANALYSIS
JOURNAL J. Phycol. 39 (5), 923-930 (2003)
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..441
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PF014h11_r"
/dev_stage="sporophytes"
/clone_lib="Porphyra yezoensis TU-1 sporophytes"
ORIGIN
Alignment Scores:
Pred. No.: 0.000559 Length: 441
Score: 138.50 Matches: 43
Percent Similarity: 43.97% Conservative: 19
Best Local Similarity: 30.50% Mismatches: 72
Query Match: 3.93% Indels: 7
DB: Gaps: 3
US-10-691-383-2 (1-676) x AU187606 (1-441)
Qy 336 ThrileAspAlaIleThrValGluProLyseGlnGluThrPheAlaProAspLeuAsnTyr 355
Db 17 ACGTACGGCGCCAAACCGGGTAGACCTGCGGGTCCCGTTCGACGCGGGGTGCGACTAC 76
Qy 356 MetValaspPheaspGluThrLeuAsnIleGlnAsnGlyGlyProProAlaGlyProGlu 375
Db 77 CTCACCGACATGGGAGTGTGTTGTTCGAGCTCCAAAACGGCGCAGACACCCCGCGCACCGGAG 136
Qy 376 GluLeu-----AspGluGluLeuArgPheIleArgAsnAlaArgAspLeu 390
Db 137 ACGTACGTGCGCGCGCGGATGGACCAAAACCCGCTTCATGTCACCCCGCGGACCTT 196
Qy 391 AlaArgValSerPheValaspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeu 410
Db 197 GCCACGTACGTCACATACGATGCTCTGTACCGGCATACCTTGAACGGGCTCTGTGGCTG 256
Qy 411 LeuGluLeuGlyAlaPheSerArgProGlyIleAsnGlyProPheIleAspSer---Asp 429
Db 257 CTAGCCGAGGTTGTGCCCTTTTGACCGCGGCTCTCCGTGGACCGCCCGCACGAGGTGGAT 316
Qy 430 ArgGlnAlaGlyPheValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAla 448
Db 317 CATCAGCAGGGGTTTGCACGTTTGGGGGGCCCCACCTGCTGACGCTCGTCACCGAGGTG 376
Qy 449 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnIleTyrGlnValHisArgPheAlaArg 468
Db 377 GCACACCGCGCACTCAGGGGGTGGCGTTCGAAAAGTTCAACGTCCACCGCCGCTCCGG 436
Qy 469 Pro 469
Db 437 CCC 439
RESULT 6
AL539002
LOCUS
DEFINITION Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS02F030Y123 5-PRIME. mRNA sequence.
linear EST 31-MAY-2003


```
Qy 348 ThrPheAlaProAspLeuAsnTyrMetValAspPhe-----AspGluTyrLeuAsn 364
Db 716 AAACCTTAACTGAT-----GGCTATGCTGTGAGTTAGATGCCAAATATCTACTTAAC 772
Qy 365 IleGlnAsnGlyGlyProProAlaGlyProGluGluLeuAspGluGluLeuArgPheIle 384
Db 773 ACAGTTATGCTGGCACAAGATGGTCCAGGATCGAGCTACAGAGCTAGCT---GTA 829
Qy 385 ArgAsnAlaArgAspLeuAlaArgValSerPheValAspAsnIleAsnThrGluAlaTyr 404
Db 830 AAATCTTGCAGAAACGCTATTAAGTTAGAGTTGTCCGGCGGAAATTCAGAGCATACATTT 889
Qy 405 ArgGlySerLeuIleLeuLeuGluLeuGlyAlaPheSer-----ArgProGlyIle 421
Db 890 GGAGAGAGACTTCGGTGTGTGCATATGGCATGTCTCTCATGATTTCTCCAGCTCTG 949
Qy 422 AsnGlyPro-----PheIleAspSerAspArgGlnAlaGlyPheVal 435
Db 950 AAGCCACCAGACGCTGGTGTCTGGTAAAGATGCTTCGGCGCAGCCCAATTTT--- 1006
Qy 436 AsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAla 455
Db 1007 AATGCTTCTGCCAAACACTGGACCAATTTTGTCT-----ATTACAGAAATGCA 1054
Qy 456 SerCysTyrGlnTyrTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThr 475
Db 1055 AAT-----GATGCAATTTGGTATCCTT 1075
Qy 476 LeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn 495
Db 1076 AACAAATTCCTCATTCACAAGATGTCATTAATGAAATATGATGATGATGCCAAAT 1135
Qy 496 -----AspGluLeuLeuLys-ArgValAlaGluIle 505
Db 1136 CGCAGATGGCGTTGTGAGTGTCTTTTACAAGATCACTGCTTAGCTGAAGGTTATGGAA-- 1193
Qy 505 eAsnAlaAlaGlnAsnProAsnGluValThrTyrLeuLeuPro-GlnAlaIleGln 524
Db 1194 -----CCAGAAACAAGTAAATCAATGAGTGTCTGAGTGTCTGAGTGTCTGAGT 1237
Qy 525 ValGlySerProThrHisProSerTyrProSer-----Gly 536
Db 1238 ATTCTTCAAAAACACAGCCCACTTATCCATCTGTCAAAAGTTCCACAATGCCATACAGGC 1297
Qy 537 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 556
Db 1298 TCTTCAACCCAGAGGATCTGGAAGAGAGAAATATAAGGATCTGTAGTTATGAGAAAT 1357
Qy 557 GlyGlyGlyCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 576
Db 1358 TCTTCA-----AATCCGTTGTCAGCTGACGAC----- 1387
Qy 577 PheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuAsnLysLeuAlaValAsnValAla 596
Db 1388 -----ACAGCTCAGTTTAAACCAATGACAGTTGATGATGTC 1423
Qy 597 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIle-----GlnGlyLeuLeu 615
Db 1424 TATGAAAGG---ATGACAGGCTCCCGTGGAAATGCAAGATGATTCTAGAGAGTGAAGTA 1480
Qy 616 LeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGlu 635
Db 1481 ATTGCAAGACAGTGTGGGTGAAG-----AAAACGTCAATATGAA 1522
Qy 636 AlaThrPheGluPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThr 655
Db 1523 GCT-----GCTGGGGAAGCTGTGAAACCTCAAAAGACCCAG 1561
Qy 656 PheSerIleAspGlyAspMetCysSerGlyLeuVal 667
Db 1562 CCAACTGTCTTAACAACCTTGAAGAAAGGAGCTGTT 1597
```

RESULT 8

```
BC040379
LOCUS
DEFINITION
Homo sapiens NF-kappa B-repressing factor, mRNA (cdna clone
IMAGE:477539), containing frame-shift errors.
ACCESSION
VERSION
BC040379.1 GI:25955669
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2673)
AUTHORS
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Dege, J.G.,
Klauser, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshikiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Adams, R.D., Mullany, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sánchez, A., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Rahvey, J., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzyzinski, M.I., Skalski, U., Smalinski, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
REFERENCE
2 (bases 1 to 2673)
AUTHORS
Strausberg, R.
DIRECT SUBMISSION
Submitted (27-NOV-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcaps-rc@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsegh, H.,
Kowals, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 29 Row: 1 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8923943
This clone has the following problem: frame shifted.
FEATURES
Location/Qualifiers
1. .2673
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:477539"
/tissue_type="Skin, normal"
/clone_lib="NCI CGAP_Skn3"
/lab_host="DH10B"
```


/note="Vector: pCMV-SPORT6.1"

ORIGIN

Alignment Scores:

Pred. No.: 4.01 Length: 2673
 Score: 115.00 Matches: 134
 Percent Similarity: 34.43% Conservatives: 96
 Best Local Similarity: 20.06% Mismatches: 260
 Query Match: 3.26% Indels: 179
 DB: 11 Gaps: 34

US-10-691-383-2 (1-676) x BC040379 (1-2673)

```

Qy 66 LeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArg 85
Db 72 CTGCGCTCGACCAAGCTCTCGCGCTCTCGCGCGCTGGACCAACCAAGCTCTCTCGGGCT 131
Qy 86 IleAlaThrAla-----Length-----LeuAlaValLeuAla 95
Db 132 GCAGGTACAGCCCAAGCTTGATGGAAAAAATTCCTCAAAATGGCTGAAGGTATTGATATTG 191
Qy 96 AlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLys 115
Db 192 GGGAGATGCCTTCA-TAATGATCGGTGCTGTCAAA-----CCTTCCAAA----- 235
Qy 116 HisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspPro 135
Db 236 -----GGTCAAAAAGCCACCTCTCAACATGT-----GATGGTCAAAATCCT 277
Qy 136 Thr-----AlaProAsnArgArgAspAsnValAlaPhe 146
Db 278 CCTAAAAAGCAAGCGGTTCCAAATTCATGCGAGACCTCGTTTGAGCCTGTACATTTT 337
Qy 147 AlaSerArgArgAspAlaAlaArgArgGluArgAspGlyThrGlyThrValCysGlnIle 166
Db 338 GTAGCTAGTAGTTCAAAAGATGAAGACAGACGAGATCCTTATGGCCCTCAACAAAAGAG 397
Qy 167 ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeu 186
Db 398 GTAATGAACAACACATTTTGCC-----AGCATGCCAAGACACATCTAC 442
Qy 187 GlyGlnValThrAlaAspAspPheAlaIleLeuGluAsp-----CysIleLeu 202
Db 443 CAAGATTATATCTCAAGACTCTTTCAAGTATCAAGATGGGAATTCAGTATTTGTGATTCA 502
Qy 203 AsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArgLeuVal 222
Db 503 TCAGGA---TTCATTCTCACAAGAACACAGCCTGTA----- 535
Qy 223 AsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr----- 240
Db 536 -----ACAGCCAACTGATTTTGTGACAGTGGGAACCTTGCCCAAGCACCATCACAG 589
Qy 241 -----ThrIleProValProThrLeuSerSerProGluLeuAla 254
Db 590 CAGGCAAACTCTCAGTCAACTCTCGAGCCTTCACCATCACAGCATTTTCCCGAG----- 643
Qy 255 AlaGlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyr 274
Db 644 TCTGTGTAGCCGAG-----AAGCAGTAT 667
Qy 275 GlyThrAspGluIleThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPhePro 294
Db 668 TTATTGAAAAAATAACGGCGACATCTGGNAGAACCTTTCTAAT-----CCA 715
Qy 295 AsnLeuAspAlaValSerIleGlySerAspGlyThrValAspPropheSerGlnLeuPhe 314
Db 716 GAAATGACTTCT-----GGATCTGATAAAAT-----AATTATACATATATGTTA 760
Qy 315 ArgAlaThrPheValGlyValGluThrGlyPro---PheValSerGlnLeuLeuValAsn 333
Db 761 ACTCGTTGTTATTCAGCGGTGTAAGACAAATCCTCGATATATATATGCTCTTTAAAGGAA 820

```

```

Qy 334 SerPheThrIleAspAlaIleThrValGluProLysGlnGluThrPheAlaProAspLeu 353
Db 821 ATTCTCTCTCGCGACATC-----CCCAAAATAAAAACTTCTAACTGAT--- 865
Qy 354 AsnTyrMetValAspPhe-----AspGluTrpLeuAsnIleGlnAsnGlyGlyPro 370
Db 866 GGCATATGCTTGTGAAGTAGATGCCAAAATATCTACTTAAGTACAGGTATGCTGGCAGC 925
Qy 371 ProAlaGlyProGluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeu 390
Db 926 AAGAATCGGTCCAGGATCGAGCTACAGAGCTAGCT---GTAAAACTCTTCAGAGAAACGT 982
Qy 391 AlaArgValSerPheValAspAsnIleAsnThrGluAlaTyrArgGlySerLeuIleLeu 410
Db 983 ATTGAAGTTAGAGTTGTCGCGCGAAATTCAGGCATACATTTTGGAGAGGACCTCGTGTGT 1042
Qy 411 LeuGluLeuGlyAlaPheSer-----ArgProGlyIleAsnGlyPro----- 424
Db 1043 TGTCAAGATTGGCATGCTCTCTATGAATTTCTCCAGCTCTGAAGCCACAGAGACCTG 1102
Qy 425 -----PheIleAspSerAspArgGlnAlaGlyPheValAsnPheGlyThrSerHis 441
Db 1103 GTGGTGTCTGGTAAAGATGCTCTGGCGACCAATTTT---AATGCTTCTGCCAACACAC 1159
Qy 442 TyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrp 461
Db 1160 TGGACCAATTTTGTG-----ATTACAGAAAATGCAAAAT----- 1192
Qy 462 GlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThrIleAla 481
Db 1193 -----GATGCAATTTGGTATCTTAAACAATTTGCTCTCATTC 1228
Qy 482 GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsn----- 495
Db 1229 AACAAAGATGTCATTTGATATACAAATATGAGATGATGCCAAATCCGACATGCGGTGTGCGA 1288
Qy 496 -----AspGluLeuLeuLys-ArgValAlaGluIleAsnAlaAlaGlnAsnPr 511
Db 1289 GTGTTTTTACAGATCACTGCTTAGCTGAAGGTTATGGAA-----CC 1330
Qy 511 AsnAsnGluValThrTyrLeuLeuPro--GlnAlaIleGlnValGlySerProThrHis 530
Db 1331 AAGAAAAACAAGTAACATGAGCTGCGCGACAGGCTTTGAAATTTCTTCAAAAACACAG 1390
Qy 531 ProSerTyrProSer-----GlyHisAlaThrGlnAsnGly 542
Db 1391 CCACATTATCACTGTCAAAAGTTCAATGCCATACAGGCTCTTCAACCAGAGGATCT 1450
Qy 543 AlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPhePro 562
Db 1451 GGAAGAAAGAAAGATATAAAGGATCTTGTAGTTTATGAGAAATTTCTTCA----- 1498
Qy 563 AsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeu 582
Db 1499 AATCCCGTGTGACCGCTGAACGAC----- 1522
Qy 583 ThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeu 602
Db 1523 -----ACACCTCAGTTTAACCGAATCAGATGATGATGTCTATGAAGG---ATGACA 1573
Qy 603 GlyIleHisTyrArgPheAspGlyIle---GlnGlyLeuLeuLeuGlyGluThrIleThr 621
Db 1574 GGCCTCGCTGGAAATGCAAAAGTATCTAGAGAGTGAAGTAAATTCGAGAGAGCAGTTGGG 1633
Qy 622 ValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAlaThrPheGluPheArg 641
Db 1634 GTGAAG-----AAAATGTCAAATATATGAAGCT----- 1660
Qy 642 LeuPheThrGlyGluValIleLysLeuPheGlnAspGlyThrPheSerIleAspGlyAsp 661
Db 1661 -----CCTGGGGAAGCTGTGAACAAACCTCAAAAGACCCCACTGCTATTAACAAAC 1714
Qy 662 MetCysSerGlyLeuVal 667

```

```

Db      1715 TTGAAGAAGGAGCTGTT 1732
      ||| |||
      : : :
RESULT 9
BI836552      889 bp      mRNA      linear      EST 04-OCT-2001
LOCUS      603089514F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5228666 5',
DEFINITION      mRNA sequence.
ACCESSION      BI836552
VERSION      BI836552.1 GI:15948102
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 889)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
      Email: cgabbs@email.nih.gov
      Tissue Procurement: Life Technologies, Inc.
      cDNA Library Preparation: Life Technologies, Inc.
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: LLAM1575 row: k column: 03
      High quality sequence stop: 808.
FEATURES             source
     1..889
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:5228666"
     /lab_host="DH10B"
     /clone_lib="NIH_MGC_120"
     /notes="Organ: pooled pancreas and spleen; Vector:
     pCMV-SPORT6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA
     source anonymous pool of spleen and pancreas from 28 yo
     male. Library is oligo-dT primed and directionally cloned
     (EcoRV site is destroyed upon cloning). Average insert
     size 1.5 kb, insert size range 1-2.5 kb. Library is
     normalized and enriched for full-length clones and was
     constructed by C. Gruber (Invitrogen). Research Genetics
     tracking code 025. Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.:      0.657      Length:      889
Score:          114.50      Matches:      66
Percent Similarity: 35.77%      Conservative: 27
Best Local Similarity: 25.38%      Mismatches: 99
Query Match:      3.25%      Indels:      69
DB:              12      Gaps:      12
US-10-691-383-2 (1-676) x BI836552 (1-889)
Qy      66 LeuProThrAspGlyIleSerAlaSerLysIleLeuGlyLysIleMetAlaArgValArg 85
      ||| : : : ||| : : :
      : : : : :
Db      57 CTGCCCTCGACGACTCTCTCGCGCTCTCCGCCGCTGACCAACCGCTCTCTCTGGGCT 116
      ||| : : : ||| : : :
      : : : : :
Qy      86 ILAAlaThrAla-----LeuAlaValValLeuAla 95
      ||| : : : ||| : : :
      : : : : :
Db      117 GCAGGTACAGCCACGCTTGATGGAAAAAATTCCTCAAAATGGCTGAAGTATTGATTTG 176
      ||| : : : ||| : : :
      : : : : :
Qy      96 AlaProCysLeuAlaPheAspGluValThrAlaSerGlyValPheProGluGluHisLys 115
      ||| : : : ||| : : :
      : : : : :
Db      177 GGAGATGCCTTCA-TATGATCTGGTCTCTCCAA-----CCTTCCAA----- 220
      ||| : : : ||| : : :
      : : : : :
Qy      116 HisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeuAspPro 135
      ||| : : : ||| : : : ||| : : :
      : : : : :

```

```

Db      221 -----GGTCAAAAACGCCACCTCTCAACATGT-----GATGCTCAAAATCCT 262
      ||| : : : ||| : : :
      : : : : :
Qy      136 Thr-----AlaProAsnArgArgAspAsnValAlaPhe 146
      ||| : : : ||| : : :
      : : : : :
Db      263 CCTAAAAAGCAACCGGTTTCCAAATTCATGCAGACCTCGTTTGTAGCCTGTACATTTT 322
      ||| : : : ||| : : :
      : : : : :
Qy      147 AlasArArgArgAspAlaAlaArgGluArgAspGlyThrGlyThrValCysGlnIle 166
      : : : : :
      : : : : :
      : : : : :
Db      323 GTAGCTAGTAGTTCAAAAGATGAAGACAGAGAGATCCTTATGCCCTCAACAAAGAG 382
      ||| : : : ||| : : :
      : : : : :
Qy      167 ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeu 186
      ||| : : : ||| : : :
      : : : : :
Db      383 GTAATGAACAAACACATTTTGGC-----AGCATGCCAAGACATCTAC 427
      ||| : : : ||| : : :
      : : : : :
Qy      187 GlyGlnValThrAlaAspAspPheAlaIleLeuGluAsp-----CysIleLeu 202
      ||| : : : ||| : : :
      : : : : :
Db      428 CAAGATTATATCAAGACTCTTTTCAGTATACAAGATGGGAATCTTCAGTATGTGATTC 487
      ||| : : : ||| : : :
      : : : : :
Qy      203 AsnGlyAspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArgLeuVal 222
      : : : : :
      : : : : :
      : : : : :
Db      488 TCAGGA---TTCATTCTCAAAAAGACCGCTGTA----- 520
      ||| : : : ||| : : :
      : : : : :
Qy      223 AsnProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThr----- 240
      ||| : : : ||| : : :
      : : : : :
Db      521 -----ACAGCCAACATGTATTTTGACAGTGGGAACCTCGCCCAAGCACCATCACAG 574
      ||| : : : ||| : : :
      : : : : :
Qy      241 -----ThrIleProProValProThrLeuSerSerProGlu----- 252
      ||| : : : ||| : : :
      : : : : :
Db      575 CAGGCAAACTCTCAGTCAACTCTCGAGCCTTCACCATCACAGACATTTCCCGAGTCTGTG 634
      ||| : : : ||| : : :
      : : : : :
Qy      253 -----LeuAlaAlaGlnLeuAlaGluLeuTyrrTrpMetAlaLeuAla 266
      ||| : : : ||| : : :
      : : : : :
Db      635 GTAGCCGAGAGACAGTATTTTATTGAAAATTAACGGGACAATCTGGAAGACCTTTCT 694
      ||| : : : ||| : : :
      : : : : :
Qy      267 ArgAspValProPheMetGlnTyrrGlyThrAspGluIleThrThrThrAlaAlaAsn 286
      ||| : : : ||| : : :
      : : : : :
Db      695 AAT-----CCAGAAATGACTCTCGGATCTGTATAAACTTACTTATACATATATGTTAAAC 748
      ||| : : : ||| : : :
      : : : : :
RESULT 10
CD063088      369 bp      mRNA      linear      EST 14-SEP-2003
LOCUS      MA1-0029U-M019-D02-U.B MA1-0029 Schistosoma mansoni cDNA clone
DEFINITION      MA1-0029U-M019-D02.B, mRNA sequence.
ACCESSION      CD063088.1 GI:34614247
VERSION      EST.
KEYWORDS      Schistosoma mansoni
SOURCE      Schistosoma mansoni
ORGANISM      Schistosoma mansoni
      Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
      Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE      1 (bases 1 to 369)
AUTHORS      Verjovski-Almeida, S., DeMarco, R., Martins, E.A.L., Guimaraes, P.E.M.,
      Ojopi, E.P.B., Paquola, A.C.M., Piazza, J.P., Nishiyama, M.F. Jr.,
      Kitajima, J.P., Adamson, R.E., Ashton, P.D., Bonaldo, M.F.,
      Coulson, P.S., Dillon, G.P., Farias, L.P., Gregorio, S.P., Ho, P.L.,
      Leite, R.A., Malaquias, L.C.C., Marques, R.C.P., Miyasato, P.A.,
      Nascimento, A.L.T.O., Ohlweiler, F.P., Reis, E.M., Ribeiro, M.A.,
      Sa, R.G., Stukart, G.C., Soares, M.B., Gargioni, C., Kawano, T.,
      Rodrigues, V., Madeira, A.M.B.N., Wilson, R.A., Menck, C.F.N.,
      Setubal, J.C., Leite, L.C.C. and Dias-Neto, E.
      Transcriptome analysis of the acelomate human parasite Schistosoma
      mansoni
      Nat. Genet. 35 (2), 148-157 (2003)
JOURNAL      22879926
MEDLINE      Contact: Dr. Sergio Verjovski-Almeida
COMMENT      Departamento de Bioquímica
      Instituto de Química - Universidade de São Paulo
      Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 São Paulo - SP,
      Brasil
      Tel: +55-11-3091-2173
      Fax: +55-11-3091-2186
      Email: verjov@ig.usp.br
      This sequence was derived from the FAPESP Schistosoma mansoni EST

```

FEATURES

```

Db      242 GTGGATCAGAACGCTCTCTAGCAGCGAATCTCTCAAGAGCTGCG-----AGA 195
Qy      453 GlnArgAlaSerCysTyrGlnLeuTrpGlnValHisArgPheAla---ArgProGluAla 471
Db      194 CAGGAAGCCCAACGACCAACGACGAGCAGCAGCAGCAATTTTCAGTATCAACCGACCG 135
Qy      472 LeuGlyGlyThrLeuHisAsn-----Thr 479
Db      134 ACTGGTCAGTCAGTGGACATCCATCCAGCCACTGCTGCCCTTCGCGCTGCAACG 75
Qy      480 IleAlaGlyAspLeu-----AspAlaAspPheAspIleSerLeuLeuGluAsn 495
Db      74 CTCACTCGGATCTGGCAATTTCCGACAGCGATGATGAGAGACCAATTTGCAGAAC 18

RESULT 12
LOCUS      CB552155
DEFINITION MMSPP0080_F05 MMSPP Macaca mulatta cDNA, mRNA sequence.
ACCESSION CB552155
VERSION    CB552155.1 GI:31301350
KEYWORDS   EST.
SOURCE     Macaca mulatta (rhesus monkey)
ORGANISM   Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
            Cercopithecinae; Macaca.
REFERENCE  1 (bases 1 to 701)
AUTHORS   Katze,M.G., Bumgarner,R., Korth,M., Feldman,R., Amjadi,M. and
            Holzman,T.
TITLE     Expressed sequence tags from Rhesus macaque spleen
JOURNAL   Unpublished (2002)
COMMENT   Contact: Holzman T
            Katze Lab
            University of Washington
            Box 358070, Seattle, WA 98195-8070, USA
            Tel: 206 732 6156
            Fax: 206 732 6055
            Email: ted@locke.hs.washington.edu
            Similar to GenBank entry HSA011812 AJ011812 Homo sapiens mRNA for
            transcription factor NRP, partial. 4/2000
            Plate: MMSPP0080 row: F column: 05.
FEATURES   source
            1..701
            /organism="Macaca mulatta"
            /mol_type="mRNA"
            /db_xref="taxon:9544"
            /sex="male"
            /cell_type="monocytes"
            /dev_stage="adult"
            /clone_lib="MMSp"
            /note="Organ: spleen"

ORIGIN
Alignment Scores:
Pred. No.:      1..15      Length:      701
Score:          110..50     Matches:      60
Percent Similarity: 39.65%  Conservative: 30
Best Local Similarity: 26.43% Mismatches: 79
Query Match:    3.13%      Indels:      59
DB:             14         Gaps:      12

US-10-691-383-2 (1-676) x CB552155 (1-701)
Qy      85 ArgIleAlaThrAlaLeuAlaValValLeuAlaProCysLeuAlaPheAspGluVal 104
Db      21 AAAATCTCCAATGCCTGAGGATTTGATATTTGGGAGATGCCCTTCA-TATGATCTGGTG 79
Qy      105 ThrAlaSerGlyValPheProGluHisIleHisThrGlyGluGlyArgCysLeuGln 124
Db      80 CTGTCCTCAAG-----CCTTCCAAA-----GGTCAAAAACGCCACTCTCA 118
Qy      125 ThrCysThrAsnSerAspAlaLeuAspProThr----- 136

```

```

Db      119 ACATGT-----GATGGTCAAAATCTCTCTAAAAAGCAAGCGGTTCCAATTC 166
Qy      137 ---AlaProAsnArgAspAsnValAlaPheAlaSerArgArgAspAlaAlaArg 155
Db      167 CATGCGAGACCTCGTTTGGCTGTACATTTTGTAGCTAGTAGTTCCAAAAGATGAAGA 226
Qy      156 GluArgAspGlyThrGlyThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThr 175
Db      227 CAGGAGATCTTATGGCCCTCAACAAAGAGAGTAAATGAACAAACACATTTTGCC--- 283
Qy      176 MetPheHisIleSerLeuProHisAspGluLeuGlnValThrAlaAspAspPheAla 195
Db      284 -----AGCTTGCCAGAGACATCTACCAAGATTATATCTCAAGACTCTTTTCAGT 331
Qy      196 IleLeuGluAsp-----CysIleLeuAsnGlyAspPheSerIleCysGluAsp 211
Db      332 ATCAAGATGGGAATTCAGTATTGTGATTCATCAGGA---TTCATCTCACAAAAGAC 388
Qy      212 ValPro-----AlaGlyAspProAlaGlyArgLeuValAsn 223
Db      389 CAGCTGTAGCAGCCACATGTATTTGACAGTGGGACCTGCG----- 433
Qy      224 ProThrAlaAlaPheAlaIleAspIleSerGlyProAlaPheSerAlaThrThrIlePro 243
Db      434 CCAAGCAGCAC-----TCACAGCAGGCAAACTCTCAGTCAACTCTCTGAG 478
Qy      244 ProValProThrLeuSerSerProGlu-----LeuAlaAla 255
Db      479 CCTTCACCATCAGACAGCATTTCCGAGTCGCTGGTAGCCGAGAGCAGTAGTATTTATTGAA 538
Qy      256 GlnLeuAlaGluLeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGly 275
Db      539 AAGTTAAGCGGCAATCTCGAAGAACCTTTCTAAT-----CCAGAGATGACTCTCTGGA 592
Qy      276 ThrAspGluIleThrThrThr 282
Db      593 TCTGATAAAATTAATTATACA 613

RESULT 13
LOCUS      AK039670
DEFINITION Mus musculus adult male spinal cord cDNA, RIKEN full-length
            enriched library, clone:A330082B18 product:hypothetical Serine-rich
            region/1PT/TIG domain/Ankyrin repeat region circular profile/Yeast
            DNA-binding domain containing protein, full insert sequence.
ACCESSION  AK039670
VERSION     AK039670.1 GI:26087328
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     92279253
PUBMED      10349636
REFERENCE   2
AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20499374
PUBMED      11042159
REFERENCE   3
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,S., Ikegami,H., Sakaguchi,S., Kashiwagi,K.,
            Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,

```

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, K., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system—384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL MEDLINE
 20530913
 11076861
 PUBMED

REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase 1 & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 2823)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saseki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Havaishizaki, Y.

TITLE Direct Submission of Genomic Sequences
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-722 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome

can initially be prepared and sequenced in mouse genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.

URL:<http://genome.gsc.riken.go.jp/>
URL:<http://fantom.gsc.riken.go.jp/>

```

FEATURES
source
1. .2823
Location/Qualifiers

```

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A330082D18"
/db_xref="MGI:2404165"
/db_xref="taxon:10090"
/clone="A330082D18"
/sex="male"
/tissue_type="spinal cord"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..2823
misc feature

```

msc_feature
1:..2523
/note="hypothetical Serine-rich region/PT/TIG
domain/Ankyrin repeat region circular profile/yeast
DNA-binding domain containing protein [PROSITE|PS50324,
InterPro|IPR002909, PROSITE|PS50297, InterPro|IPR003163,
evidence: InterPro]"

ORIGIN				evidence: INCEIFLO)
Alignment Scores:				
Pred. No.:	13.2	Length:		2823
Score:	110.50	Matches:		132
Percent Similarity:	34.74%	Conservative:		74
Best Local Similarity:	22.26%	Mismatches:		202

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match:	3.13%	Indels:	188
DB:	11	Gaps:	30
US-10-691-383-2 (1-676) x AK039670 (1-2823)			
Qy	162	ThrValCysGlnIleThrAsnGlyGluThrAspLeuAlaThrMetPheHisLeuSerLeu	181
Db	1167	ACTTCCTGTGAGCTTGCCGCTGGTGTAGGTT-CTGAGCTTCTGCCCCACCGGCTCCTC	1225
Qy	182	Pro-----	184
Db	1226	CGAGCCCTGCCTTGACCCCTGATCGTTTTCTCAACAGCCCCCAAGGGCCAGACATACG	1285
Qy	185	GluLeuGlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGly	204
Db	1286	GAGGG-GGCCAAGAGTAAACCCAGACTTCCCGAGCAGAG-----GGC	1329
Qy	205	AspPheSerIleCysGluAspValProAlaGlyAspProAlaGlyArgLeuValAsnPro	224
Db	1330	ACTCACACTCCCTGT-----CCTGCCTCGAACCTGCAGCTGCCCTGGAGCCCCAG	1380
Qy	225	ThrAlaAlaPheAlaIle-----	230
Db	1381	CGGCTGCTCGAGGCTCTCGCTCCACAGTGGGAGCAACGGAGAGAGAGAAACAAATTC	1440
Qy	231	-----AspIleSerGlyProAlaPheSerAlaThrThr	241
Db	1441	TTTATCCAAGATGATGATAGTGGGGAGAACTCAAGGGTCCGGGA-----ACG	1488
Qy	242	IleProProValProThrLeu-----SerSerProGluLeuAlaAlaGlnLeuAlaGlu	259
Db	1489	GTGCGCGCGTACTTCTATCCCTCTCTTATCCCGCTCTCCCAACCGCTTGCGCCCA	1548
Qy	260	LeuTyrTrpMetAlaLeuAlaArgAspValProPheMetGlnTyrGlyThrAspGluIle	279
Db	1548	-----	1548
Qy	280	ThrThrThrAlaAlaAlaAsnLeuAlaGlyMetGlyGlyPheProAsnLeuAspAlaVal	299
Db	1549	TCAGGCGAGGCAACGAGAGGAGAAAGCCTTGTTCGAGGC-----	1587
Qy	300	SerIleGlySerAspGlyThrValAspProPheSerGlnLeuPheArgAlaThrPheVal	319
Db	1588	TCGTCTGGCAGCAGCAGCTCGAGCCTTCAGT-----	1623
Qy	320	GlyValGluThrGlyProPheValSerGlnLeuValAsnSerPheThrIleAspAla	339
Db	1624	---CTTTCGTCAATCCACAGACCTCATGGGAGAACTCATCATGAT-----GAAGCT	1671
Qy	340	IleThrValGluProIlysGlnGluThrPheAlaProAspLeuAsnTyrMetValAspPhe	359
Db	1672	CCAGGTGTCCCTGCCCGCCCCCGACCTCTCTCTCGCTAAATGCATCAGACACTTC	1731
Qy	360	Asp---GluTrpLeuAsnIleGlnAsnGlyGlyProPro-----AlaGlyPro---	374
Db	1732	TCCCCAGATGG---TCCTACCCAGAGGGCGGGTCAGGTGCTCATCAGAGCCCTTGG	1788
Qy	375	GluGluLeuAspGluGluLeuArgPheIleArgAsnAlaArgAspLeuAlaArgValSer	394
Db	1789	ACAGAGCCCGCAGGCAATTACTCTGTGTC-----	1818
Qy	395	PheValAspAsnIleAsnThrGluAlaTyrArgGlySerIleIleLeuLeuGluLeuGly	414
Db	1819	---TTCGATCACATC-----GCAGTCCAGCCTCCCTGTGTC-----	1851
Qy	415	AlaPheSerArgProGlyIleAsnGlyProPheIleAspSerAspArgGlnAlaGlyPhe	434
Db	1952	-----CAGCTGGTGTCTTACGCTGCTACTGTCTCTGCC---CATGAGGTAGGGCTG	1899
Qy	435	ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg	454
Db	1900	GTGTCTTGG-----CAGGTGGCGGGCGAGAGGCGCCCTCTCT	1938

Qy 455 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 474
 Db 1939 GCCTCTGCTCTTGTAGTATCGAGCCGCGGTCTCTG-----TCTCTGCCTAGT 1989
 Qy 475 ThrLeuHisAsnThrIleAlaGlyAspLeuAspPheAspIleSerLeuLeuGlu 494
 Db 1990 ACACAGCTCGACTGTTGTCTCACTGGAT---GACAGCCAGTTCGGATGTCATCTCGAG 2046
 Qy 495 Asn---AspGluLeuLeuLysArgValAlaGluIleAsnAlaLa----- 508
 Db 2047 CGCTGAGCAGATGGAGAGCGGATGCGAGATTCCGGCAGCTGGCGGCTCTCGGT 2106
 Qy 509 -----GlnAsnProAsnAsnGlu----- 514
 Db 2107 CAGGCCCGCAGGCTCTCCATTCAGATGAGGCCAGGCGCTCTGCTCGAGGACGG 2166
 Qy 515 -----ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHis 530
 Db 2167 GTGTGTGCTCTGTAGAGCAGCATGATCCCGGTCCACTTG-GAGGGGTCTGAACGGCT 2225
 Qy 531 ProSerTyrProSer-----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 548
 Db 2226 GATCATCGGAAGTCCCTTCGCGGCGCATGAGCTCTACATCTGCTGCTGCACAGGGCTA 2285
 Qy 549 LysAlaLeu-----IleGlyLeu 554
 Db 2286 CGCTCGGCTCATCGAGACTCTGAGCCAGTGGAGGAGTGTGGAAACAGGAGCTTGACTT 2345
 Qy 555 AspArgGlyGly-----GluCysPheProAsnProValPheProSerAspAspGly 571
 Db 2346 AGAGCAAGAGGTTGACCGCTCAATGTGGACCATTTCTTGTGACCCCTCTGATGTGGC 2405
 Qy 572 Leu-GluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLe 591
 Db 2406 CTGTGCTCTGGACACATTAGAAGCTGTGCTCTTTCTTTTGTGGAAACGACAGCACT 2465
 Qy 591 uAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIle-----HisTy 606
 Db 2466 GAGCATTCGGAGTCTCTGGCGGAGTCCCTCTGTGGCTCATCTCTGCTGCTCAGT 2525
 Qy 606 rArgPheAspGlyIleGlnGlyLeuLeuGlyGluThrIleThrValArgThrLeuHi 626
 Db 2526 CGCGCTTCGCGCTGCTCTGAGGAACCTGACAGACAGGAGCTTTCAGTTGAG-----CA 2579
 Qy 626 sGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGl 646
 Db 2580 CCCACGCGCTCTATCC---CCACATCCTCTAGCCCGACACTGGCCCTGAGACGCGCTC 2636
 Qy 646 uValIleLysLeuPheGlnAspGlyThrPheSerIle 658
 Db 2637 GTACCCCTCGAGCTGCATGTCGCACTTTCTCTGTG 2673

RESULT 14
 CNS03HC2/c
 LOCUS
 DEFINITION
 1052 bp DNA linear GSS 01-SEP-2000
 Tetraodon nigroviridis genome survey sequence T7 end of clone
 026D21 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION
 AL244091
 VERSION
 AL244091.1 GI:7965103
 KEYWORDS
 GSS; genome survey sequence.
 SOURCE
 Tetraodon nigroviridis
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
 1
 Roest Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,
 Bernot.A., Fzames.C., Wincker.P., Brottier.P., Quetier.F.,
 Saurin.W. and Weissenbach.J.
 Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)
 20296633
 10835645
 REFERENCE
 2
 Roest Crolius.H., Jaillon.O., Dasilva.C., Ozouf-Costaz.C.,
 Fzames.C., Fischer.C., Bouneau.L., Billault.A., Quetier.F.,
 Saurin.W., Bernot.A. and Weissenbach.J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Genome Res. 10 (7), 939-949 (2000)
 20359837
 10899143
 REFERENCE
 3 (bases 1 to 1052)
 Direct Submission
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 1. 1052
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="026D21"
 /clone_lib="G"
 /notes="Genoscope sequence ID : COBG026CB11LP1-end : T7"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.64 Length: 1052
 Score: 110.00 Matches: 54
 Percent Similarity: 36.36% Conservative: 18
 Best Local Similarity: 27.27% Mismatches: 86
 Query Match: 3.12% Indels: 40
 DB: 29 Gaps: 9
 US-10-691-383-2 (1-676) x CNS03HC2 (1-1052)
 Qy 71 IleSerAlaSerIlybilleLeuGlyLysIleMetAlaArgValArgIle-----Ala 87
 Db 1046 GTCCCCACGTCGAGGCGGCGGAGGCCATCGCGAGCAGCGGTCCAGACCATGCAR 987
 Qy 88 ThrAlaLeuAlaValValLeuAlaAlaProCysLeuAlaPheAspGluValThrAlaSer 107
 Db 986 ACCACGCGCGCACGCTCCMAGAAGCTCATACGCTTGGCGTAGAAGCGGTGGTGGCGCAT 927
 Qy 108 Gly-----ValPheProGluGlu 113
 Db 926 GGCCATGTAGCGGTGACGCTGATGAGAGACATGAGGCGAGCGTGGAAACAGAGAG 867
 Qy 114 HisLysHisThrGlyGluGlyArgHisLeuGlnThrCysThrAsnSerAspAlaLeu 133
 Db 866 CACGCGCAT-----GAGGCCACACCTTGCA-----GCTCAGGACGCTGTA 825
 Qy 134 AspPro-----ThrAlaProAsnArgAspAsnValAlaPheAlaSerArg 149
 Db 824 GTTCCCGAGCGGAGCGCTTCCGTACCGCACACAGGACGAGGAGAGCGMCCGA 765
 Qy 150 ArgAspAlaAlaArgArgGluArgAspGlyThr-----GlyThrValCysGlnIle 166
 Db 764 GCGAGCGCTGTCGCCGAGCACAGTCCAGCAGGAAGTAGTAGGCGCTTGTGAGCGCG 705
 Qy 167 ThrAsnGlyGluThrAspLeuAlaThrMetPheHisLysSerLeuProHisAspGluLeu 186
 Db 704 CGGTCTCTGAGGACGAGCGGACAC-----AGCAGGTTGCCACCCAGTGCACG 654
 Qy 187 GlyGlnValThrAlaAspAspPheAlaIleLeuGluAspCysIleLeuAsnGlyAspPhe 206
 Db 653 CAGATGATGATCCCGAGCAGCAGCTTGTACGCTGGCGGAGACGCGCGCGG----- 600

This Page Blank (uspto)

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	127.5	3.6	650	2	H84356	long-chain fatty-a
2	120	3.4	1742	2	S76110	hypothetical prote
3	118.5	3.4	2254	2	D86215	protein 16D2.14 [
4	116.5	3.3	145	2	F75447	vanadium chlorop
5	115	3.3	750	2	A84315	malate dehydrogen
6	113.5	3.2	912	1	IJCHCN	N-cadherin precurs
7	113.5	3.2	3472	2	T31308	hypothetical 367K
8	113	3.2	1374	2	AE3259	extracellular seri
9	112.5	3.2	7576	2	T17428	FK506 polyketide s
10	112	3.2	1019	2	AB3613	conserved hypothet
11	111	3.1	876	2	E96674	hypothetical prote
12	110.5	3.1	439	2	B44315	carthage oligomer
13	110.5	3.1	504	2	AF1789	gluconate kinase h
14	110.5	3.1	794	2	C87437	penicillin-binding
15	109.5	3.1	1965	2	S75200	fat protein - Syne
16	108.5	3.1	165	2	C70888	hypothetical prote
17	107.5	3.0	646	2	G83006	esterase EstA PA51
18	107.5	3.0	797	2	AE6232	succinoglycan bios
19	107.5	3.0	797	2	AD1054	hypothetical prote
20	107.5	3.0	3624	2	AD0835	large repetitive p
21	107	3.0	2273	2	T09083	hemagglutinin/hemo
22	107	3.0	3972	2	S75251	.hypothetical prote
23	107	3.0	5188	2	B85547	probable RTX fami
24	106.5	3.0	947	2	H86933	probable DNA topoi
25	106.5	3.0	1389	2	I50090	carboxypeptidase 9
26	106	3.0	426	2	H69127	histidinol dehydro
27	106	3.0	1121	2	T06065	hypothetical prote
28	106	3.0	1122	2	T14180	exit protein - Myc
29	106	3.0	3705	2	AD0123	probable autotrans

Db 534 VAQAMVVGDGQKFSAIL--VPNFDVSEWAAQSEIALP-----DDRAICRDERVRAR 585
Qy 393 V-SPVDNINT 401
Db 586 IQSAVDDVNT 595
RESULT 2
576110
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76110
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76110
A>Status: nucleic acid sequence. not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1742 <KAN>
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10088.1; PID:d101073
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG
Query Match 3.4%; Score 120; DB 2; Length 1742;
Best Local Similarity 19.4%; Pred. No. 2.5;
Matches 153; Conservative 77; Mismatches 242; Indels 316; Gaps 39;
Qy 107 SGVPEEHKHTGEGRHLOCTC---NSDDA----LDPTAPN--RRDNVAFASRRDAARRE 156
Db 684 NGVVTGNPANTGDCGTITATNAGNTNDVQGEIDTTSPTGTPNGNSPF-SPLDITDPT 742
Qy 157 RDGTGTCQITNGTDLATMFHKSPLHDELG---QVTADEFAILED----- 199
Db 743 DTGADDLS--SDGNPELTFTGEGPCLEILLGSDGNSVDPDAVELETTPGVNPGDPSTYTI 801
Qy 200 CILNGDFSICEDVPAGD-----PAGR-----LNPTAAFAIDISGPA 236
Db 802 KLIDADSPDSDFGDFNGSPGNGANTGDGIYITATDNAGTNDVQGEID----- 856
Qy 237 FSATTTPPVPTLSPPELAQAELAEYMWALARDVPFMQGTDEITTTAAANLAGMGGPPNL 296
Db 857 ---TTSPTGTPNGNS-----PFSPLDITDATDTGADLLSGNGNPFL 895
Qy 297 -----DAVSGSDGTVDPP 310
Db 896 TFGESGLEILLGPDGPNVDPDAVEVETPGANPGDPSTYTIKLVDADPPDSDSDPP 955
Qy 311 SQLFRATFVG--VETGPFVSQLLVNSFTIDA-----ITVEPKQET 348
Db 956 GDFPENGVTGNPANTGD-----GTITATDNAGTNDVQGEIKTSSLSINDRGT 1008
Qy 349 FA-PDLNY-----MVPDFEWNLTQ-----NGGPPAGPEE-----LDRE 380
Db 1009 VWIGDLGFGSINVDLANDQGLGIQITNLQPLKGTAVLRDNGTTPNNFSDFLTYIHEE 1068
Qy 381 LRFRNARDLARVSFVNINTEAVRGSLSL-----ILLELGAFSRP----- 419
Db 1069 LLDURSATGVRIDISIFSSASVNSVNGFYQVLDHFGTILDPASGRLLTPADLEYRNA 1128
Qy 420 -----GINGPFTSDRQAGFVNFCTSHYFRILIGAELAQRA 455
Db 1129 LASAIPGLQISKONPVQSIELPGGIVAPFLTSDQNGQVN-----HFFALIDA----- 1177
Qy 456 SCYQKQVH--RPARPALGGT--LHNTIAGDLAD---FDISLENDEL-----LKR 502
Db 1178 ---NPGQVDYVREA-----NGTYSFEDSAFGDRFNDLRAVNILPGQGFVPPAIGLTAN 1229

Qy 503 AEINAAQNPNEVITYLLPQ-----AIOQVSPTHPSPSGHATQNGAFATVLKALIGL 554
Db 1230 AGVFTANGLVDQFTYITDVNGDNRTASVAIDNQHLKF-----TLGV 1273
Qy 555 DRGEGCFPNPV--FPSDDGLEINFEGACLTYSGEINKLANVA--FGRQMLGIHYRFDG 610
Db 1274 D--GD-FKNEVAIFRVDD-----RQGTINGIAPGQAGVLEAALSNSQVIFTG 1317
Qy 611 IQG--LLLGETITVRIHQELMTFAEATPEFLFTGEVILKFDQGTFSIDGDMCSGLVY 668
Db 1318 ISGQTQFGESLT-RTMEE---FKVSDSFGFLIQQNNSLDM-----VLRQAAGAIS 1365
Qy 669 TGVAQCQA 676
Db 1366 TPVFFSQA 1373
RESULT 3
D86215
protein T6D22.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86215
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Schartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86215
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2254 <STO>
A:Cross-references: GB:AE005172; NID:g8778840; PIDN:AAF79839.1; GSPDB:GN00141
C:Genetics:
A:Gene: T6D22.14
A:Map position: 1
Query Match 3.4%; Score 118.5; DB 2; Length 2254;
Best Local Similarity 20.7%; Pred. No. 4.8;
Matches 107; Conservative 68; Mismatches 174; Indels 167; Gaps 26;
Qy 134 DPTAPNRRDNVAFASRRDAARRE-----RDGTGTCQITNGTDLATMFHKSPLH 183
Db 1723 DPAANTESENLEAIEPQASAGSETVETDFAASHQGDQVTCPLSSPTG---NQPAPE 1777
Qy 184 DEL-GQ--VTADDFAILEDCLINGDFSCIEDVPAGD-----PAGLVNPTAAFAIDIS 233
Db 1778 ANTEGONINTASPHVAGPDVAVESGDYAVIDQBTMGQAQACSLPSSG-VGTQSDLGANIE 1836
Qy 234 GFAFSATTIPPVPTLSSPELAA-----OLAEYMWALARDVPFMQGTDEITTTAAAN 286
Db 1837 GQ--NVTVAQLFDGSDAVVTCGSPVSDCAQ-----DASPMPLSPGHNHPT---AVN 1886
Qy 287 LAGMGFPNLDVAVSI-----GSDG---TVDPFSQLFRATFV-----QVETGPFVSQ 330
Db 1887 IEG-----LDNTSVAPHHISGSDACEMEISEPGQOVERSTFANLPHFEGVHSAGVTAL 1940
Qy 331 ----LVNSFTIDATVPEPKQETFPADLNMYVDPEWLNQNGGPPAGPEELDELRFINA 387
Db 1941 VPSLLANGT-EQIAVOP-----VPQIPFPVFNDFPLH-----ELEKURE 1979
Qy 388 RDLARVSFVDNINTEAVRGSLSLLELCAFRSGINGPFFIDSDRQAGFVNFCTSHYFRILG 447
Db 1980 SENSKTPEKVS-----FHYPKSIL 2001
Qy 448 AAELAQRASCYQKQVHRFARPEALGGTTLHNTIAGDLADFDI-----SLLEN 495

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83613
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1019 <STO>
A;Cross-references: GB:AE004464; GB:AE004091; NID:9946099; PIDN:AAG03651.1; GSPDB:GN0011
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA0262

```
Query Match      3.2%; Score 112; DB 2; Length 1019;
Best Local Similarity 20.3%; Pred. No. 4.5;
Matches 140; Conservative 81; Mismatches 255; Indels 212; Gaps 35;

Qy 52 IKYLEG-----SETWKKVKLPDGTGSAKILGKIMAR-----VRIATALAVL 94
Db 436 VDVLGDCDQPIVGTGRYATNRPYALPDHKILSTIKSKYKGSRANELRIDDTTAQIS 495
Qy 95 AAPCLAFDEVTASGVF-----PEEHKHTGEGRHLOTCTNSDDALDPTAPNRDNVAF 146
Db 496 AA-----LMSDHGASALHGLYTHPRPEGKPRGEGFELT-----DEHGVRAAKGLL 544
Qy 147 AS-----RRDAARRERDGTGVCQITNGETDLATMF-----HKSLPHDELQGVTTADFFA 195
Db 545 LSTEEQLRAGAGHLDR---GVVVQVLEALELARELGDYAGEHQGVGHDAAPQOT-----596
Qy 196 ILEDCTINGDFSCEDV---PAGDPAGRLVNPTAFADISGPAPSAITIPVPVTLSPPEL 253
Db 597 -LOEAV-----RDLGHGANDBSGKNGKPAIA--LSGPAGIAATAATPASLTAAAGEH 645
Qy 254 AAQLAELVWALARDVPFMYQGTDEITTTAAANLAGMGFPNLDVAGISGDTVPDSQL 313
Db 646 VDSVARONQQTAGQKVVINAGSD-----IGLFAQGG-----EL 679
Qy 314 FRATFVGVETGPFVFSQLLVNSFTIDAITVBPQK--ETFPADLNVMVDFDEWLNIQNGGPP 371
Db 680 RQIT---HQPMLLOAKND---IRLEAKQSVESASQOHVLTVAKEHITLMCGG-- 728
Qy 372 AGPEELDEELFRINARDLARVSVDNINTEAY---RGSLLILELGAFSRPGINGPIDS 428
Db 729 -----AYLTGKGNII--ELG-----MPG-----744
Qy 429 DRQAGFNVFGTSHYFRILGIAELAAQRAQSCYQKQVHRFARPEALGG-----TLH 477
Db 745 ----NFVVKAAKSH---VGAASLEAELPQFVETQRRFVLKQDGTAMPNPVYTITMA 798
Qy 478 N--TIAGDLADFDISLENDELKRVAEINAAQNPNNEVTYL---LPQAIQVGSPT-HP 531
Db 799 NGEVIEGVTDAEGATQLLOKDAMNIAKVDKMTKSPASAVAGIAAAGAAVAVGKLLGGP 858
Qy 532 SYPSGHATQNGAFATVULKALIGDRGCECPNVPFSPDDGLELINF-EGACLTVEGINK 590
Db 859 DAEAGRALSEGE-----ISLAG-----VFGDSIDYSTVRLRDEDEYVPMQGDYV 903
Qy 591 LAVN--VAFRGQMLGI-HYRFDGI--QGLLL-----GSTITVRLTHOSLMTFAE 634
Db 904 MAPNGHIYFGEELRGVADWSLESQROGLFIHETHVWQHGVNVLVGVAYQAROFLL 963
Qy 635 EATPEFRLFTGEVILKFDQGTFSIDGDM 662
Db 964 GDQVAYRLEPKTLK---DYNIEQQGDI 988
```

RESULT 11
E96674
hypothetical protein F16G16.7 [imported] - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A;Accession: E96674
R;Theoretical: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
anssen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, B.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E96674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-876 <STO>
A;Cross-references: GB:AE005173; NID:g6227005; PIDN:AAF06041.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16G16.7
A;Map position: 1

```
Query Match      3.1%; Score 111; DB 2; Length 876;
Best Local Similarity 20.3%; Pred. No. 4.2;
Matches 105; Conservative 68; Mismatches 184; Indels 160; Gaps 23;

Qy 221 LVNPTAFAIDISGPAPSAITIPVPTLSP-----ELAAQLAELVW-----263
Db 59 LCNQLSPFASTTNG--LSATKNAEIPVGNSPESRNLNLSAALAAEMWKKRGLGLSE 116
Qy 264 -----ALARDVPFMYQGTDEITTTAAANLAGMGFPNLDVAGISGDTVPDSQLFRAT 317
Db 117 IQDLSDIVERAVSGQLTVRELCTVSTLTAASTQKLKAAI--SDNRVTFPLVDILQ-- 173
Qy 318 FVGVEGTFVFSQLLVNSFTIDAITVBPQKQETFPADLNVMVDFDEWLNIQNGGPPAGPEEL 377
Db 174 --GDFKDTLQKKI--SFCIDC-----NMTMILD-----RA 200
Qy 378 DEELRFRNARDLARVSVDNINTEAYRGSLLILELGAFSRPGINGPIDSROAGFVNF 437
Db 201 SEDLEIRSR-----RRNMENLDSLKKIKTIFLAGINKPLITQRSRMCVAI 251
Qy 438 GTSHYFRILGAAEL---AQRASCYQKQVHRFARP-EALGCTLHN-----T 479
Db 252 RATHSKLLPGVVLSVSSSPATC-----FIEPREAV--ELNNMEVRHANSEKAEEMA 301
Qy 480 IAGDLADFDISLENDELKRVAEINAA-----QNPNNEVTYLFPQAIQV 525
Db 302 ILSILTSEVVMQAQREILHLDRILELDAFARASHANWINGVYPNVTSEHTKTPGLAVDI 361
Qy 526 GSPTHSYPSGHATQNGAFATVULKALIGDRGCECPNVP-FPSDDGLELINFEG-----579
Db 362 DSAQHP-----LLLSVLGSPNGGDIFFVPVDIKVESAKVVVISGPNVTGG 407
Qy 580 --ACLTVEGINKLVNVAFRQWMLGIHYR-----FDGIQGLLGLGETITVRLTHQE 628
Db 408 KTAALLKTLGLLSMSKS-----GNVLPKNCPLPWFED---LILADIGDPQSLEQS 455
Qy 629 LMTFAEATPEFRLFTGEVILKFDQGTFSIDGDMCSG 665
Db 456 LSTFSGHIS--RI--RQILDIASENSLVLDBICSG 487
```

RESULT 12
B44315
cartilage oligomeric matrix protein - bovine (fragment)
C;Species: *Bos primigenius taurus* (cattle)
C;Date: 09-Dec-1993 #sequence_revision 03-May-1994 #text_change 23-Dec-2002
C;Accession: B44315; S36710
R;Oldberg, A.; Antonsson, P.; Lindblom, K.; Heinegard, D.
J. Biol. Chem. 267, 22346-22350, 1992
A;Title: COMP (cartilage oligomeric matrix protein) is structurally related to the throm-
A;Reference number: A44315; MUID:93054522; PMID:1429587
A;Accession: B44315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-439 <OLD>
A;Cross-references: EMBL:X74326; NID:g396625; PIDN:CAA52374.1; PID:g396626

Db 846 SGDTAKAIAENSGANQVIYQVTATDDADVSGITF 880

Search completed: September 19, 2004, 03:59:48
Job time : 44.9931 secs

RESULT 15

S75200
fat protein - *Synechocystis* sp. (strain PCC 6803)
N;Alternate names: protein slr2046
C;Species: *Synechocystis* sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75200
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, I.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75200
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1965 <XAN>
A;Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BAAL17114.1; PID:dl01784
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	3.1%;	Score 109.5;	DB 2;	Length 1965;
Best Local Similarity	21.0%;	Pred. No. 19;		
Matches 146;	Conservative 78;	Mismatches 250;	Indels 221;	Gaps 36;
Qy	97	PCLADEVETASGVPEEHKHTGEGRHQLQCTNSDADLPTAPNRDRNVAFASRRDAAARRE	156	
Db	284	PSVNNNVTV-GFFPFGSQGVNLISQTSILTYGTPPTDNTFPNA-----PSTPDLSSASS	336	
Qy	157	RDGTGTVCCQINGT-----DLATMPHKSPLHDELGOVTAD---DFAILEDCLN	203	
Db	337	DSGLSSDNIINDTPTFNGTAEANSTVTLF--SGGSTQIGSTANGSNWTTIATSPAD	394	
Qy	204	GDESICEDVPAGDPAGRL-----VNPTAAFAIDISGPAF-----	237	
Db	395	GNYSI--TAKATDAAGNVSTASSALGITINDTNTPLNLSAIEISDTALKIGTATVTTFTFS	452	
Qy	238	-----SATTIPPVPTLSSPELAQLAELYMALARDVPPMQYGTDEIT--TTAAANLA	288	
Db	453	EAVIGFTNADIIVDGSLSSP--TSSDGGITWTATLTPNAAESNVITLDTNGTISDLA	510	
Qy	289	GMGGF-----PNLDAVSTGSDG-----TVDPF	310	
Db	511	GNNGTGTTTSVSYAVDTPPTLTISIDGDDDNIVPIDTPLYTLTFSSEIDSTTVTADDF	570	
Qy	311	SQLFRAFF-VG--VETGPFVSQLLANSTIDAIVTEPKQETFPD-----LNNYVDEDEW	362	
Db	571	DNAGTAISIGTITETSSGVFSVVPTWTSGTIILOIPNGAVLSDMAGNNLAVPQDDDE	630	
Qy	363	LNIQNGGPPA--GPEELDEELRFIRNARDLARVSVDNINTEAYRGSLLLELGA--FSR	418	
Db	631	LOV-NOGFSAVIVPNASLAENTDTNPLKVADIAITDD-----GLGSNDISL	676	
Qy	419	PGINGPIDSDRQAGFNVFGTSHYFRLIGAELAQRASCYQKWQVHREAPAEALGGLIHN	478	
Db	677	SGIDAAAFEVIGQSFLTKAGTVLNF-----SKAS-----YTVNVVDDTVGST---	721	
Qy	479	TIAGDLADDFDI SLLLENDEL-----LKRVAEINAAQPNNEVTVLLPQAIVQSPT	529	
Db	722	---PDLTADFKLTINNLDEVAPITISGDTAKAIAENSGA---NQIIV-----QV---	764	
Qy	530	HPSVPSGHATONGAFATVLIKALIGLDRGCEFPNPVPSDDGLE-LINFEAGCLTYEGEI	588	
Db	765	-----KATDGDGISA-----GITFG-----LKPFGDADSFIIN-----ATTVEG	798	
Qy	589	NKLAVNVAFGRQMLGIHYRF-----DGTQGLLGETITVTRLHOELMTFAEATFEPLF	643	
Db	799	-KLIGNDDPFETQS---SYKFTVTADGGV-----NPATEQLVNLAIIDLADIAP---	845	
Qy	644	TGEVIKLFQDG-----TFSIDGDMCSGLV	668	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 03:56:03 ; Search time 254.636 Seconds
(without alignments)
3303.314 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFQSHYFRLGAEALQRA.....GLLGGTITVRLHQLMTF 198

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10691383/runat_17092004_102655_1733/app_query.fasta_1.1230
-DB=N_Geneseq_29Jan04 -QFW=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691383@cgn 1.1 653 @runat_17092004_102655_1733 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002s:*
- 7: Geneseq2003as:*
- 8: Geneseq2003bs:*
- 9: Geneseq2003cs:*
- 10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	2931	3 AAA10305	AAA10305 Fucus van
2	1039	100.0	2931	6 AAD28155	AAD28155 Fucus dis
3	332.5	32.0	1797	4 AAD11050	AAD11050 Corallina
4	329.5	31.7	1794	2 AAV56020	AAV56020 Haloperox
5	322	31.0	1791	2 AAV56021	AAV56021 Haloperox
C 6	108	10.4	110000	4 AAI99682_42	Continuation (43 o
C 7	108	10.4	110000	4 AAI99683_42	Continuation (43 o
8	101.5	9.8	597	7 ACA53911	ACA53911 Prokaryot

9	101	9.7	5828	3 AAA14845	AAA14845 DNA encod
10	99	9.5	51	3 AAA10306	AAA10306 Curvulari
11	99	9.5	51	6 AAD28156	AAD28156 Probe #1
12	91.5	8.8	1161	4 AAH22872	AAH22872 R. ruber
13	91.5	8.8	10480	4 AAH22877	AAH22877 R. ruber
14	88.5	8.5	1191	4 ABL28263	ABL28263 Drosophil
15	88.5	8.5	2117	4 AAF81756	AAF81756 Human mem
16	88.5	8.5	2122	4 AAD12569	AAD12569 Human pro
17	88.5	8.5	2192	9 ADE31718	ADE31718 Human 646
C 18	88	8.5	819	7 ACA42710	ACA42710 Prokaryot
19	87.5	8.4	1815	6 AAH77240	AAH77240 DNA of ha
20	87.5	8.4	1815	6 ABA03638	ABA03638 Dreschler
C 21	87.5	8.4	1706	9 ADB69915	ADB69915 C. neofor
22	87	8.4	1812	6 AAS15767	AAS15767 D. salina
23	86	8.3	579	7 ACA39594	ACA39594 Prokaryot
24	86	8.3	3058	5 AAS06054	AAS06054 Angiotens
C 25	86	8.3	4334	5 AAS06053	AAS06053 Angiotens
C 26	85.5	8.2	1830	2 AAT00803	AAT00803 Chloroper
27	85.5	8.2	2088	2 AAT03875	AAT03875 Chloroper
28	85.5	8.2	2763	6 ABA92252	ABA92252 Mouse Her
30	85.5	8.2	2781	6 ABA92253	ABA92253 Her-2/neu
31	85.5	8.2	3771	3 AAB89737	AAB89737 Mouse Her
32	85.5	8.2	3771	5 AHA42210	AHA42210 Nucleotid
33	85.5	8.2	3771	6 ABA92251	ABA92251 Mouse Her
34	85.5	8.2	6273	7 ACC44344	ACC44344 Gene enco
35	85.5	8.2	12279	6 ABQ93367	ABQ93367 Human cDN
36	84.5	8.1	6197	6 ABS78947	ABS78947 E. coli C
37	84	8.1	504	5 AAH68141	AAH68141 C glutami
C 38	84	8.1	309400	5 AAH68534	AAH68534 C glutami
39	83.5	8.0	525	7 ACA28777	ACA28777 Prokaryot
40	83.5	8.0	1161	4 AAH22892	AAH22892 R. ruber
41	83.5	8.0	12127	4 AAD05497	AAD05497 Human sec
42	83.5	8.0	2228	6 ABQ99407	ABQ99407 Human cod
43	83.5	8.0	2317	6 ABS76414	ABS76414 cDNA enco
44	83.5	8.0	2317	6 ABV99386	ABV99386 Human NOV
45	83.5	8.0	2822	2 AAT51612	AAT51612 Curvulari

ALIGNMENTS

RESULT 1

AAA10305

ID AAA10305 standard; cDNA; 2931 BP.

XX

AC AAA10305;

XX

DT 03-JUL-2000 (first entry)

XX

XX Fucus vanadium bromoperoxidase cDNA.

DE Vanadium bromoperoxidase; vanadium haloperoxidase; alga; ODA;

XX O-dianisidine oxidation; vanadium ion; halogenation; antimicrobial agent;

KW phenolic adhesive production; ss.

XX Fucus sp.

OS

XX Key Location/Qualifiers

PH CDS 228..2258

FT /*tag= a

FT /product= "Fucus vanadium bromoperoxidase"

XX WO200015771-A1.

XX PD 23-MAR-2000.

XX 27-AUG-1999; 99WO-US019893.

XX 10-SEP-1998; 98US-00151189.

XX (REGC) UNIV CALIFORNIA.

XX Vreeland V, Ng KL;

```
XX WPI: 2000-271414/23.
DR P-PSDB; AAY87469.
XX
PT Fucus vanadium bromoperoxidase polypeptide and coding sequence, useful
PT for halogenating or oxidizing compounds, e.g. as antibacterial agents.
XX
PS Claim 3; Page 36-41; 46pp; English.
XX
CC This sequence represents cDNA encoding a vanadium bromoperoxidase from
CC species of Fucus. This enzyme is 73.4 kD in mass and catalyses the
CC oxidation of o-dianisidine (ODA) when complexed with a vanadium ion. The
CC cDNA encoding the enzyme was isolated from a Fucus 2-cell embryo cDNA
CC library using hybridisation probes (AA10306-A10307) corresponding to the
CC second and third regions that are conserved between Curvularia and
CC Ascophyllum vanadium haloperoxidase active sites. Fucus vanadium
CC bromoperoxidase cDNAs (one full-length, two 5'-truncated) were then
CC cloned into the pET LIC (ligation independent cloning) vector for
CC expression in Escherichia coli. The vanadium bromoperoxidase proteins
CC were expressed fused to an N-terminal thioetheroxin tag which optimises
CC correct folding. The Fucus vanadium bromoperoxidase has a specific
CC activity several-fold higher than other algal vanadium haloperoxidases
CC for which at least partial sequences have been reported. Vanadium
CC haloperoxidases can be used in a method for enzymatically halogenating a
CC compound, and in a method for enzymatically oxidising a compound. They
CC can be used to halogenate various substances, including proteins. They
CC can be used to produce epoxides from alkenes, halogenated ketones from
CC alkenes, to produce alpha, gamma-halohydrins from cyclopropanes, and to
CC produce dihalogenated products from alkenes and alkynes. Vanadium
CC haloperoxidases may also be used to oxidise various compounds, making
CC them useful in signal generating systems in place of horseradish
CC peroxidase. The enzymes can also be used as enzymatic antimicrobial
CC agents and in the production of phenolic adhesives
XX
SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.66e-113 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAA10305 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTTCTCAGATTGATAGTGCCTGGAGCTGGCGCAGCGT 1589

Qy 21 AlaSerCysTyrGlnTyrGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGATCGATTTTCACGCCGCCGAGGCTCTCGGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACACACCATCGCGGGGATCTAGATGCAGACTTCGACATCTCCCTCTTGAA 1709

Qy 61 AsnAspGluLeuLeuIleArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGTGGCGAGATAATTCGGCGCAGAGATCCCAACACGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTACCTACCTTCTTCACAGCTATCCAAAGTGGGATCGGACAGCCCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyValaPheAlaThrValLeuIleAlaLeuGlyLeu 120
Db 1930 TCCGCCACGCTACCCAAATGGAGCATTTGCCACAGTCTTGAGAGGCCCTCATTTGGCCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTGTCTTCCCTTAACCCCGTGTTCACAGCGATGATGCGGCTGGAACCTA 1949
```

```
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGGATGCCCTTACATATAGGAGAGATCAACAGCTCGGGTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCAATTTGGAGGAGGATGCTGGGCATCCATTCGTTTCGACGGTATCCAAAGGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACAATCACTGTACGAACAATTCACAGGAGCTGATGACGTTTC 2123

RESULT 2
ID AAD28155 standard; cDNA; 2931 BP.
XX
AC AAD28155;
DT 07-AUG-2003 (revised)
DT 01-MAY-2002 (first entry)
XX
DE Fucus distiches. vanadium haloperoxidase encoding cDNA.
XX
KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
KW industrial catalysis; enzyme; catalyst; ss.
XX
OS Fucus distichus.
XX
FH Key Location/Qualifiers
FT CDS 228..2258
FT /tag= a
FT /product= "vanadium haloperoxidase"
XX
PN WO200200838-A2.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US018602.
XX
PR 19-JUN-2000; 2000US-00596794.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Vreeland V;
XX
DR WPI; 2002-154651/20.
DR P-PSDB; AAE17501.
XX
PT Isolated nucleic acid for, e.g. to halogenate various substrates
PT including proteins, has polynucleotide sequence encoding vanadium
PT haloperoxidase polypeptide with catalytic helical frame.
XX
PS Claim 4; Page 49-53; 56pp; English.
XX
CC The invention relates to nucleic acid encoding vanadium haloperoxidase
CC particularly vanadium bromoperoxidase polypeptide which consists of a
CC catalytic helical frame that complexes a vanadium ion and catalyses the
CC oxidation of o-dianisidine (ODA). The invention also provides
CC recombinantly produced vanadium haloperoxidases. The polypeptides of the
CC invention can be fused to other proteins to allow quantification or
CC localisation of the linked protein. They can be used to halogenate
CC various substrates including proteins; to produce epoxides from alkenes,
CC halogenated ketones from alkynes; to produce alpha, gamma-halohydrins
CC from cyclopropanes and to produce dihalogenated products from alkenes and
CC alkynes. They can also be used in signal generating systems in place of
CC horseradish peroxidase; as a component in assays; as enzymatic
CC antimicrobial agents; and in the production of phenolic adhesives. The
CC polypeptides can be used in industrial catalysis in a variety of
CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
CC reactions. The present sequence is Fucus distiches vanadium
CC haloperoxidase encoding cDNA. (Updated on 07-AUG-2003 to correct OS
```

```

CC field.)
XX SQ Sequence 2931 BP; 694 A; 776 C; 784 G; 677 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 9.66e-113 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAD28155 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGCCGCCGAGCTGGCGACGGT 1589
Qy 21 AlaSerCysTyrGlnIlystTglnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTTACCAAAAGTGGCAGGTGCATCGATTTCACGCCGCCGAGGCTCTCGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACACACCATCGCGGGGATCTAGATCGACATTCGACATCTCCCTTCITGAA 1709
Qy 61 AsnAspGluLeuLeuIlystArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTGAAACGTGTGGCGGAGATATAATGGCGGAGATCCACACGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTTACCTTCTTCACCAAGCTATCCAAAGTGGATCGCAACGCCCTTCCTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIlystAlaLeuIleGlyLeu 120
Db 1830 TCGGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTGTCTTCCCTAACCCGTGTCCACAGCATGACGCGCTGGAATA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnIlystLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGCATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCATTTGGGAGGAGATGCTGGGCATCCACTATCGGTTCCGACGATATCCAGGCGCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCCTGGAGAGACAAATCACTGTACGAAACACTTCACGAGGCTGATGAGCTTC 2123

RESULT 3
AAD11050
ID AAD11050 standard; cDNA; 1797 BP.
XX AC AAD11050;
XX DT 24-SEP-2001 (first entry)
XX DE Corallina officinalis vanadium bromoperoxidase (cVBPO) cDNA.
XX KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
XX KW signal generation; analytical system; anti-microbial agent;
XX KW industrial application; biological material; avidin; metabolite;
XX KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
XX KW vitamin; intact cell; drug; narcotic; BS.
XX OS Corallina officinalis.
XX FH Key Location/Qualifiers

CDS
FT 1..1797
FT /*tag= a
FT /product= "Corallina officinalis vanadium bromoperoxidase
FT (VBPO)"
FT primer_bind
FT complement(1..21)
FT /*tag= b
FT /bound_moiety= "PCR forward primer (AAD11054)"
FT primer_bind
FT 1150..1170
FT /*tag= c
FT /bound_moiety= "PCR reverse primer (AAD11055)"
FT misc_binding
FT complement(1420..1449)
FT /*tag= d
FT /bound_moiety= "PCR probe (AAD11053)"
FT primer_bind
FT complement(1420..1442)
FT /*tag= e
FT /bound_moiety= "PCR upstream primer (AAD11051)"
FT primer_bind
FT 1642..1664
FT /*tag= f
FT /bound_moiety= "PCR downstream primer (AAD11052)"
XX WO200153494-A2.
XX 26-JUL-2001.
XX 24-JAN-2001; 2001WO-US002383.
XX 24-JAN-2000; 2000US-0177764P.
XX (REGC ) UNIV CALIFORNIA.
XX Butler A, Carter JN;
XX WPI; 2001-457611/49.
XX P-PSDB; AAE05750.
XX Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
XX obtained from Corallina officinalis for use in signal generation in
XX analytical systems.
XX Claim 25; Fig 1; 68pp; English.
XX The present sequence is Corallina officinalis (marine red algae) vanadium
XX bromoperoxidase (cVBPO) cDNA. The vanadium haloperoxidase, especially
XX vanadium bromoperoxidase (VBPO) is useful for signal generation in
XX analytical systems, and as anti-microbial agents. The VBPO is also useful
XX for industrial applications. The VBPO is also useful for detecting a wide
XX variety of chemical and biological materials such as amino acids,
XX peptides, polypeptides, proteins including enzymes, avidin, antibodies
XX and antigenic proteins, carbohydrates including monosaccharides,
XX polysaccharides and lipopolysaccharides, hormones such as human chorionic
XX gonadotropin, thyroid stimulating hormone, luteinising hormone,
XX thyroxine, follicle stimulating hormone, parathyroid hormone and growth
XX hormone, metabolites such as glucose, lactate and pyruvate.
XX oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
XX cells from various organisms including microorganisms and drugs such as
XX narcotics, therapeutic and those abused
XX SQ Sequence 1797 BP; 394 A; 480 C; 513 G; 410 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.8e-29 Length: 1797
Score: 332.50 Matches: 82
Percent Similarity: 56.34% Conservative: 38
Best Local Similarity: 38.50% Mismatches: 72
Query Match: 32.00% Indels: 21
DB: 4 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x AAD11050 (1-1797)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 19
Db 1120 GTGAACCTTCGGAGACGACACGCTGCTGAGTCTGTGTGACTGAGGTGCTACGCGCGCTTG 1179

```

Qy 20 ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 39
Db 1180 AAGCGGTACGTTACCAAGAGTTTAACTTCATCTCGCTCGGCTGAGGCTACCGGT 1239
Qy 40 GlyThrLeuHis---AsnThrIleAlaGlyAsp-----LeuAspAlaAspPhe 54
Db 1240 GGTCTGATTAGCGTGAACAAATCGCAGCGAGAGCGGCGAGAGCGTTTCCCTGAGGTT 1299
Qy 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAala 74
Db 1300 GACCTCTCTGTTGAAGAGCTTGAAGATATCTCGAGAAAGCTGAAATTAGCAATAGAAA 1359
Qy 75 GluAsn-----ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 89
Db 1360 CAGAACAATGCTGACGAGATCTTGACCTGATCTTCATCTCTGTTGCCGAGGCATTC 1419
Qy 90 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyVala 109
Db 1420 GCCGAGGCGAGCCCATTCATCCGTCCTACGGAAGCGCCACGCTGTGTGTTGCTGGCGCA 1479
Qy 110 PheAlaThrValLeuLysAlaLeuIleGly-----LeuAspArgGlyGlyGlu 125
Db 1480 TGTGTGACGATCTCAAGGCGTTCTTCGACTCCAACTTCAGATCCGATCGATGTTCCAG 1539
Qy 126 CysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsnPheGluGly 145
Db 1540 -----GTCGACAAGATGAGGACAAAGCTTGTAAAGTCTCTTTCAAGGGA 1584
Qy 146 AlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 165
Db 1585 ACT---CTCACTGTTGCGGTGAGTGAACAAGCTCGCCGACAAATATTGCGATCGGCGG 1641
Qy 166 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 185
Db 1642 AACATGCGAGGTGTCACTACTTCTCTGACCAAGTTCGAGTCAATCTGCTTGTGTGAGCAG 1701
Qy 186 IleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 1702 GTTGGATGGATCTTGGAGAGCAGACAGTCTGACGTAT 1740

RESULT 4

AAV56020

ID AAV56020 standard; cDNA; 1794 BP.

XX

AC AAV56020;

XX

DT 09-DEC-1998 (first entry)

XX

DE Haloperoxidase, CP.BP01 encoding cDNA.

XX

KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;

KW bromine; iodine; ds.

XX

XX Corallina pilulifera.

XX

PH Key Location/Qualifiers

FT CDS 1..1794

FT /*tag= a

FT /product= "Haloperoxidase, CP.BP01"

FT /note= "the stop codon is not indicated"

XX

XX JP10248581-A.

XX

XX 22-SEP-1998.

XX

XX 06-MAR-1997; 97JP-00070539.

XX

XX 06-MAR-1997; 97JP-00070539.

XX

XX (SAKA) OTSUKA PHARM CO LTD.

XX

XX WPI; 1998-560733/48.

XX

XX P-PSDB; AAW80550.

XX

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

XX

SQ

New haloperoxidase gene - and corresponding vector, transformed host cell and method of preparation.

Claim 2; Page 13-15; 20pp; Japanese.

This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina pilulifera. A host cell transformed with a vector containing the haloperoxidase genes can be used for the recombinant production of the enzyme. The haloperoxidases of the invention are vanadium-dependent and have halide specificity of bromine and iodine

Sequence 1794 BP; 395 A; 481 C; 505 G; 413 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.61e-29 Length: 1794
Score: 329.50 Matches: 85
Percent Similarity: 56.13% Conservative: 34
Best Local Similarity: 40.09% Mismatches: 74
Query Match: 31.71% Indels: 19
DB: 2 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x AAV56020 (1-1794)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 19

Db 1120 GTGAATCTCGATCCGACACACGTGTGAGTCTGACTGAGTGTGCTACGCGCGCTTG 1179

Qy 20 ArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGly 39

Db 1180 AAGCGGTACGTTACCAAGAGTTTAACTTCATCTCGCTCGGCTGAGGCTACCGGT 1239

Qy 40 Gly---ThrLeuHisAlaLeuIleAala-----GlyAspLeuAspAlaAspPhe 54

Db 1240 GGTCTGATTATCTTAAACAAATCGCACCGCAGAGGCGAGAGCATTTTCCCTGAGGTT 1299

Qy 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAala 74

Db 1300 GATCTTGTCTGTTGAAGAGCTTGAGAGATATCTTGAGAAAGCTGAATAGCAATAGGAAG 1359

Qy 75 GlnAsn-----ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIle 89

Db 1360 CAGACATAGCTGACGAGATCTGACCTGATCCTTCATCTGTTGCGGATGCAATC 1419

Qy 90 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 109

Db 1420 GCCGAGGCGAGCCCATTTCCATCCGTCCTACGGAAGCGCCACGCTGTGTGTTGCTGGCGCA 1479

Qy 110 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsn 129

Db 1480 TGTGTGACGATCTCGAAGGCTTC-----TTCGATCCGCGCATCGAGATC-----GAT 1527

Qy 130 ProValPheProSerAsp-----AspGlyLeuGluLeuIleAsnPheGluGlyAla 146

Db 1528 CAGGTGTTGAGGTCGACAAAGATGAGGACAAGCTTGTGAAGTCGCTTTCAAGGGAAT 1587

Qy 147 CysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 166

Db 1588 ---CTCACTGTTGCGGTGAATTTGAACAAGCTCGCGACAATATTGCGATCGGCGGTAAC 1644

Qy 167 MetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGluThrIle 186

Db 1645 ATGGCAGGTGTTCACTACTTCTCTGACAGTTCAGTTCAGTCTTGTCTCGTGTGAGCGTT 1704

Qy 187 ThrValArgThrLeuHisGlnGluLeuMetThrPhe 198

Db 1705 GCGATTGGAATCTTGAAGAGCAAAAGTCTGACGTAT 1740

RESULT 5

AAV56021

ID AAV56021 standard; cDNA; 1791 BP.

XX


```
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Alignment Scores:
Pred. No.: 5.26 Length: 110000
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservatives: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72
DB: 4 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x AAI99682_42 (1-110000)

Qy 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db 70277 GGACTCTCCCACTTCGCGAGCACTGCATCGTTGGCTGATCTGCGCTGCTGGCGCG 70218

Qy 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 70217 ATAGCCCTGCGACGGCGCGC-----CGGAATGGCTTGTG-----70182

Qy 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db 70181 ---CGCGGGCGCGCGCTTCGTCGCCCATGCCATCGCC-----70146

Qy 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAala 74
Db 70145 -----GTCTGATCAAGCGGTGTGGCGG-----CGT 70119

Qy 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db 70118 CAGCGGCGGATCATCCGGCCATC-----CGGTCAACGTGGACACGCCA 70074

Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db 70073 AGTCACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGCGCTGCTCATG 70014

Qy 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Db 70013 GGCAGAGCCACCGGCTG-----CCGTACCGGTGTG-----69981

Qy 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db 69981 -----69981

Qy 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Db 69980 -----CTGGTGGCGCGGATGGCGCTGCGGAATACTGTCGGGGTCCACTACCCAGT 69927

Qy 175 AspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 186
Db 69926 GATGTGGCGGTGGGTGTGTTCTGTCGGCGCCACCGTC 69891

RESULT 7
AAI99683_42/c
Continuation (43 of 44) of AAI99683 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
```

```
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Alignment Scores:
Pred. No.: 5.26 Length: 110000
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservatives: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72
DB: 4 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x AAI99683_42 (1-110000)

Qy 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db 62534 GGACTCTCCCACTTCGCGAGCACTGCATCGTTGGCTGATCTGCGCTGCTGGCGCG 62475

Qy 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 62474 ATAGCCCTGCGACGGCGCGC-----CGGAATGGCTTGTG-----62439

Qy 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db 62438 ---CGCGGGCGCGCGGTTCGTCGCCCATGCCATCGCC-----62403

Qy 55 AspIleSerLeuLeuGluAsnAspGluLeuLysArgValAlaGluIleAsnAlaAala 74
Db 62402 -----GTCTGATCAAGCGGTGTGGCGG-----CGT 62376

Qy 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db 62375 CAGCGGCGGATCATCCGGCCATC-----CGGTCAACGTGGACACGCCA 62331

Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db 62330 AGTCACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGCGCTGCTCATG 62271

Qy 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Db 62270 GGCAGAGCCACCGGCTG-----CCGTACCGGTGTG-----62238

Qy 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db 62238 -----62238

Qy 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
```

Db 62237 -----CTGGTCCGCGATGCGCTTCGGAATACCTGCTGGGGTCCACTACCCAGT 62184
Qy 175 AspGlylleGlnGlyLeuLeuGlyGlyClnrHrile 186
Db 62183 GATGTGCGCGTGGGTGTCTCTGGGCGCCACCGTC 62148

RESULT 8
ID ACA53911 standard; DNA; 597 BP.
XX AC ACA53911;
XX AC ACA53911;
XX 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #35568.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Yersinia pestis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABUS0041.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 41781; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 597 BP; 115 A; 144 C; 138 G; 200 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0.0221 Length: 597
Score: 101.50 Matches: 39
Percent Similarity: 42.86% Conservative: 18
Best Local Similarity: 29.32% Mismatches: 46
Query Match: 9.77% Indels: 30
DB: 7 Gaps: 5

US-10-691-383-2_COPY_435_632 (1-198) x ACA53911 (1-597)

Qy 77 ProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHis 96
Db 87 CCTGATTATGCTCATCCCTATTTTACTGATAGGGCTGTGGCTATGGGGCCAAAAGATAC 146
Qy 97 ProSerTyrProSerGlyHisAlaThrGlnAsnGly----- 108
Db 147 -----CATGGACATGCAACGTACTGTGTGTCAGCAAGACAGCTATCGCTCTCGC 194
Qy 109 -----AlaPhe-AlaThrValLeuLysAlaLeuIleGlyLeuAspArgGly----- 123
Db 195 TTTTTCATGCTTTCTGCTACCTGATGCGGATGCTCATTCCTCATGATCGGCTTTGT 254
Qy 124 -----GlyGluCysPhe-----ProAsnProValPheProSerAspAs 136
Db 255 CGATGGCTTTGGTTATGCTTTATGAGCCATGCGCCAGACAGCTCTTTCCAGTGACCA 314
Qy 136 pGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGly----- 152
Db 315 CGGACCGCTATTATTTCATCTTTTGCACCTGCGGCTTTCTTTTGGCATCACCTTTGGTCCGG 374
Qy 153 ----GluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHi 171
Db 375 TGTGAGCCTGATGTTTATGCCATGCGATGCTGCTGTCAGTATCTATCTGCGGTGTCA 434

171 sTyrArgPheAspGlyIleGlnGlyLeuLeuGly 183
Db 435 TTGCGCATTCGACATGCTGGTGGTGTCTGTAGGT 471

RESULT 9
ID AAA14845 standard; DNA; 5828 BP.
XX AC AAA14845;
XX 08-AUG-2000 (first entry)
XX DE DNA encoding toluene monooxygenase proteins.
XX KW Toluene monooxygenase; TomK; TomL; TomM; TomN; TomO; TomP; TomQ; Toluene;
XX ortho-cresol; 3-methylcatechol; chlorinated compound;
XX aliphatic hydrocarbon; trichloroethylene; dichloroethylene; phenol;
XX aromatic compound; benzene; cresol; environmental remediation; ss.
XX OS Burkholderia cepacia.
XX PH Key Location/Qualifiers
XX CDS 216..446
XX FT /*tag= a
XX FT /product= "TomK protein"
XX FT /note= "encodes AAY84816"
XX FT 391..1458
XX FT /*tag= b
XX FT /product= "TomL protein"
XX FT /note= "encodes AAY84817"

CC are conserved between *Curvularia* and *Ascophyllium* vanadium haloperoxidase
 CC active sites. Fucus vanadium bromoperoxidase cDNAs (one full-length, two
 CC 5'-truncated) were then cloned into the pET LIC (ligation independent
 CC cloning) vector for expression in *Escherichia coli*. The vanadium
 CC bromoperoxidase proteins were expressed fused to an N-terminal
 CC thiorodoxin tag which optimises correct folding. The Fucus vanadium
 CC bromoperoxidase has a specific activity several-fold higher than other
 CC algal vanadium haloperoxidases for which at least partial sequences have
 CC been reported. Vanadium haloperoxidases can be used in a method for
 CC enzymatically halogenating a compound, and in a method for enzymatically
 CC oxidising a compound. They can be used to halogenate various substances,
 CC including proteins. They can be used to produce epoxides from alkenes,
 CC halogenated ketones from alkynes, to produce alpha, gamma-halohydrins
 CC from cyclopropanes, and to produce dihalogenated products from alkenes
 CC and alkynes. Vanadium haloperoxidases may also be used to oxidise various
 CC compounds, making them useful in signal generating systems in place of
 CC horseradish peroxidase. The enzymes can also be used as enzymatic
 CC antimicrobial agents and in the production of phenolic adhesives. The
 CC present sequence represents a hybridisation probe corresponding to the
 CC second region of conservation between the active sites of *Curvularia*
 CC *inaequalis* vanadium chloroperoxidase and *Ascophyllium* vanadium
 CC bromoperoxidase, which was used in the isolation of cDNA encoding Fucus
 CC vanadium bromoperoxidase. (Updated on 06-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 51 BP; 12 A; 21 C; 9 G; 9 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.00143 Length: 51
 Score: 99.00 Matches: 17
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.53% Indels: 0
 DB: 3 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAA10306 (1-51)

Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110
 Db 1 CCAACGCACCGCTTCGTACCGCTCTGGCCAGCTACCCAAACGAGCATTT 51

RESULT 11
 AAD28156
 ID AAD28156 standard; DNA; 51 BP.
 XX
 AC AAD28156;
 XX
 DT 01-MAY-2002 (first entry)
 XX
 DE Probe #1 used for cloning Fucus *distichus* vanadium haloperoxidase gene.
 XX
 KW Vanadium haloperoxidase; vanadium bromoperoxidase; O-dianisidine; ODA;
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
 KW industrial catalysis; catalyst; enzyme; probe; ss.
 XX
 OS *Ascophyllium* sp.
 OS *Curvularia* *inaequalis*.
 XX
 PN WO200200838-A2.
 XX
 PD 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018602.
 XX
 PF 19-JUN-2000; 2000US-00596794.
 XX
 PR (REGC) UNIV CALIFORNIA.
 XX
 PA Vreeland V;
 XX
 PI WPI; 2002-154651/20.
 XX
 DR Isolated nucleic acid for, e.g. to halogenate various substrates
 XX including proteins, has polynucleotide sequence encoding vanadium
 PT

PT haloperoxidase polypeptide with catalytic helical frame.
 XX
 PS Example 1; Page 28; 56pp; English.

XX The invention relates to nucleic acid encoding vanadium haloperoxidase
 CC particularly vanadium bromoperoxidase polypeptide which consists of a
 CC catalytic helical frame that complexes a vanadium ion and catalyses the
 CC oxidation of O-dianisidine (ODA). The invention also provides
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the
 CC invention can be fused to other proteins to allow quantification or
 CC localisation of the linked protein. They can be used to halogenate
 CC various substrates including proteins; to produce epoxides from alkenes,
 CC halogenated ketones from alkynes; to produce alpha, gamma-halohydrins
 CC from cyclopropanes and to produce dihalogenated products from alkenes and
 CC alkynes. They can also be used in signal generating systems in place of
 CC horseradish peroxidase; as a component in assays; as enzymatic
 CC antimicrobial agents; and in the production of phenolic adhesives. The
 CC polypeptides can be used in industrial catalysis in a variety of
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
 CC reactions. The present sequence is a probe used for cloning Fucus
 CC *distichus* vanadium bromoperoxidase gene. This probe is designed based on
 CC the second conserved regions between *Ascophyllium* sp. and *Curvularia*
 CC *inaequalis* vanadium peroxidase active site

SQ Sequence 51 BP; 12 A; 21 C; 9 G; 9 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.00143 Length: 51
 Score: 99.00 Matches: 17
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.53% Indels: 0
 DB: 6 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AAD28156 (1-51)

Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110
 Db 1 CCAACGCACCGCTTCGTACCGCTCTGGCCAGCTACCCAAACGAGCATTT 51

RESULT 12
 AAD282872
 ID AAD282872 standard; DNA; 1161 BP.
 XX
 AC AAD282872;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE R. ruber cddB gene ORF encoding lauryl lactone esterase enzyme.
 XX
 KW Cyclododecanone; degradation; lauryl lactone esterase; cdda; cddB; cddx;
 KW dodecanolic diacid; cyclododecanone monooxygenase; cyclic ketone; cddy;
 KW 12-hydroxylauric acid dehydrogenase; 12-Oxo lauric acid dehydrogenase;
 KW macrolactone; microbiological; cddC; cddD; ds.
 XX
 OS *Rhodococcus* ruber.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1161
 FT /*tag= a
 FT /gene= "cddB"
 FT /product= " lauryl lactone esterase"

WO200142436-A2.
 PN 14-JUN-2001.
 XX
 PD 08-DEC-2000; 2000WO-US033426.
 XX
 PF 10-DEC-1999; 99US-0170214P.
 XX
 PR (DUFO) DU PONT DE NEMOURS & CO E I.
 XX

Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservative: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: 4 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x AAF22877 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnValThrTyrLeuLeuValHisArgPheAla 33
Db 2951 GCCGGGGTTCGGCGCGCTACCAACGCAACACCGGCGCCACCTCATCGAGGAAT 3010
Qy 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisValThr-IleAlaGlyAs 49
Db 3011 CGTCCCGCGCATCACCGCGCGAGTTCAAGGAGTTTCGTCACACACATCGCGCGTCC 3070
Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLeuValAl 69
Db 3071 GCTCGCTGCCACTTCAGATCGCGCGCGGAGAACGAT-----TGGGGCGCTACCGC 3124
Qy 69 aGluileAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
Db 3125 CGAGATCGTCGCAACCCCTTCGACATCGATCTCGCGCACTCGATCCGACAGCGT 3184
Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
Db 3185 CATGGTGAAGACGATGACGGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 3244
Qy 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuVal 116
Db 3245 CGCGCAGCGGACATGGGAGCCCTCAACGGTCAAGGCAACGCGGCGGCGGCGGCGG 3304
Qy 116 aLeuileGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 136
Db 3305 CCTCTGACCATCAGCTGGCGGTGAG-----AGCAA 3337
Qy 136 pGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsn-- 155
Db 3338 CGGCGCTCGCTCTCCACAGCAGCCTCGATCGGCAAGATTTTCGAGAGCAGAACACGA 3397
Qy 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172
Db 3398 CGTGCACCTCGTCTCGGTGTGGCTTCGCGCGC-----GGCATCGGCTAC 3443

RESULT 14
ABL28263
ID ABL28263 standard; DNA; 1191 BP.
XX AC ABL28263;
XX 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36262.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEXE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 36262; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fip.wipo.int/pub/published_pct_sequences
XX Sequence 1191 BP; 281 A; 312 C; 326 G; 272 T; 0 U; 0 Other;
SQ

Alignment Scores: 2.01 Length: 1191
Pred. No.: 88.50 Matches: 43
Score: 38.89% Conservative: 20
Percent Similarity: 26.54% Mismatches: 60
Best Local Similarity: 8.52% Indels: 39
Query Match: 4 Gaps: 6
DB:

US-10-691-383-2_COPY_435_632 (1-198) x ABL28263 (1-1191)

Qy 48 GlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 67
Db 193 GGTAAATTTGGATGCCCTCTCTCGACGCACTTCTGCGAACGAACTTTTGGCAACCGG 252
Qy 68 ValAlaGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGln 87
Db 253 GTC---GAATTTATCCCAACAGCTGCTCAATCTGCAGCGCGCGACCTGCTCCCGCT 309
Qy 88 AlaileGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsn 107
Db 310 GGCAGGGCGATGGATCCCATACATCCGACGTGGGCCCAATCGGTGCTCACCTTCGAC 369
Qy 108 GlyAlaPheAlaThrValLeuLysAlaLeu----- 117
Db 370 GGCGCGGAGCAATGTGATGAAGACCTTCTGCGAGGTGCACAATTTGCACCTGAGGTC 429
Qy 118 -----IleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSerAsp 135
Db 430 GAGCGCTATGAGCGCGACAACTCGGGAGGAATTTACGACAA-----GAATCAGC 480
Qy 136 AspGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsn 155
Db 481 GATGGCATG-----CTGGGGGATATCTAC 504
Qy 156 LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuHisTyrArgPheAsp 175
Db 505 GAGCAGCGCGTGGAGATGGCCATTGA-----TGTATCTACAATTTGATGAT 552
Qy 176 GlyileGlnLeuLeuLeuGlyGluThrIleThrValargThrLeuHisGlnGluLeu 195
Db 553 GGAATC-----ACGAAACCTCGCACCATAGCCCGTCTTCG 591
Qy 196 MetThr 197
Db 592 GTGACC 597

RESULT 15
AAF81756
ID AAF81756 standard; cDNA; 2117 BP.
XX AC AAF81756;
XX 12-JUN-2001 (first entry)
XX DT

XX DE Human membrane associated protein MEMAP-16 encoding cDNA.

XX KW Human; membrane associated protein; MEMAP; diagnosis; cytostatic;

XX KW antiinflammatory; anticonvulsant; immunosuppressive; antidiarrhetic;

XX KW antiarteriosclerotic; gene therapy; cell proliferative disorder;

XX KW autoimmune disorder; inflammatory disorder; neurological disorder;

XX KW gastrointestinal disorder; cancer; inflammation; atherosclerosis;

XX KW epilepsy; diarrhoea; BS.

XX OS Homo sapiens.

XX PN WO200112662-A2.

XX PD 22-FEB-2001.

XX PF 14-AUG-2000; 2000WO-US022315.

XX PR 17-AUG-1999; 99US-0149641P.

XX PR 09-NOV-1999; 99US-0164203P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Lal P, Yue H, Tang YT, Bandman O, Burford N, Azimzai Y;

XX PI Baughn MR, Lu DAM, Patterson C;

XX DR WPI; 2001-168860/17.

XX DR P-PSDB; AAB74710.

XX PT Isolated polypeptide with a human membrane associated protein sequence is

XX PT useful for the diagnosis, prevention and treatment of cell proliferative,

XX PT autoimmune/inflammatory, neurological and gastrointestinal disorders.

XX PS Claim 5; Page 159-160; 173pp; English.

XX CC AAF81741 to AAF81777 encode the human membrane associated proteins

XX CC (MEMAP) given in AAB74695 to AAB74731. MEMAPs have cytostatic,

XX CC antiinflammatory, anticonvulsant, immunosuppressive, antidiarrhetic and

XX CC antiarteriosclerotic activities, which can be used in gene therapy.

XX CC MEMAPs and agonist of MEMAPs can be used to treat a disease or condition

XX CC associated with decreased expression of functional MEMAP and antagonists

XX CC of MEMAP are used to treat a disease or condition associated with

XX CC overexpression of functional MEMAP. These disorders include cell

XX CC proliferative, autoimmune/inflammatory, neurological and gastrointestinal

XX CC disorders. The MEMAP polynucleotides and proteins are also used for the

XX CC diagnosis of these disorders. Specific examples of these disorders

XX CC include cancer, inflammation, atherosclerosis, epilepsy and diarrhoea.

XX CC MEMAP proteins can be used to screen for compounds which specifically

XX CC bind MEMAP including antibodies, oligonucleotides, proteins and small

XX CC molecules. MEMAP polynucleotides can be used to prepare transgenic

XX CC animals which can be studied to provide information concerning human

XX CC disease. Anti-MEMAP antibodies are useful in immunoassays for the

XX CC detection of MEMAP protein and can be used as antagonists to treat or

XX CC prevent a disorder associated with MEMAP. Polynucleotides encoding MEMAP

XX CC can be delivered to target cells with genetic abnormalities with respect

XX CC to the expression of MEMAP to treat or prevent a disorder associated with

XX CC MEMAP

XX SQ Sequence 2117 BP; 339 A; 731 C; 672 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.48	Length:	2117
Score:	88.50	Matches:	38
Percent Similarity:	41.79%	Conservative:	18
Best Local Similarity:	28.36%	Mismatches:	39
Query Match:	8.52%	Indels:	39
DB:	4	Gaps:	7

US-10-691-383-2_COPY_435_632 (1-198) x AAF81756 (1-2117)

Qy 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnValHisArgPheAlaArg 34

Db 320 GCCCGCTCAGTGGCGCGGCTCTGTAC-----CTCAGCAAC 358

Qy	35	ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGly-----AspLeuAspAla	52
Db	359	CCCGAG-----GGCACCTGTGAGGACACTCGGGCTGGCCTCTGGGCTCTCATGCA	409
Qy	53	AspPheAspIleSerLeuLeuGluAsnAspGlu-----Leu	64
Db	410	GACCACCTCTGGCCCTGTCTGAGAGGCCCAAGGCTGACCCCGGGCTGTGAGTGGCTG	469
Qy	65	LeuIleArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsn-----GluVal	81
Db	470	CTGCAGAGGATGCAGGCCCGGGCTGCGGCCAGACCCCAAGACGGCTCGTAGATATC	529
Qy	82	ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSer	101
Db	530	CCTCAGCTGTGAGGAGCGGTGGGGCGGGGCTCCG-----	568
Qy	102	GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIleGlyLeuAsp	121
Db	569	-----GGCAGTGTGGGGGCTCTGGCTGCCCTGTGGACCATGTC	610
Qy	122	ArgGlyGlyGluCysPhe-----ProAsnProValPhe	132
Db	611	AGGAGCGGGTCTTGTCTTCCACGCTTCCGAGCCCTCAGTAC	652

Search completed: September 19, 2004, 04:27:09
Job time : 303.636 secs


```
; OTHER INFORMATION: vanadium bromoperoxidase
US-09-840-762A-1

Alignment Scores:
Pred. No.: 1,06e-134 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x US-09-840-762A-1 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGGCCGCGAGCTGGCGACGCT 1589
Qy 21 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaAspPheAspIleSerLeuGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCACAGCCCGGAGGCTCTCGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuGlu 60
Db 1650 ACCCTCCACAACACCACCGGGGATCTAGATGCAGCTTCGACATCTCCCTTTTGAA 1709
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGGCGGAGATAAATGGCGGAGATCCCAACAGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTACCTCTTCCACAAGCTATCCAAAGTGGATCGCCACGACCCCTTCTACCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCGGGCCAGCTACCCAAAATGGAGCATTTGCCACAGTCTGAAGGCGCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGTCAAC 2009
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTGCGATTGGGAGCGAGATGCTGGGCATCCACTATCGGTTCGACGGTATCCAAGGCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACATCACTGTACGAACACTTCACCGAGGCTGATGAGCTTC 2123

RESULT 2
US-10-691-383-1
; Sequence 1, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR FILING DATE: 2003-06-19
; PRIOR FILING DATE: 2000-06-19
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2931

; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-10-691-383-1

Alignment Scores:
Pred. No.: 1,06e-134 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x US-10-691-383-1 (1-2931)
Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGGCCGCGAGCTGGCGACGCT 1589
Qy 21 AlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaAspPheAspIleSerLeuGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCACAGCCCGGAGGCTCTCGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuGlu 60
Db 1650 ACCCTCCACAACACCACCGGGGATCTAGATGCAGCTTCGACATCTCCCTTTTGAA 1709
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGGCGGAGATAAATGGCGGAGATCCCAACAGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACTACCTCTTCCACAAGCTATCCAAAGTGGATCGCCACGACCCCTTCTACCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCGGGCCAGCTACCCAAAATGGAGCATTTGCCACAGTCTGAAGGCGCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGTCAAC 2009
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTGCGATTGGGAGCGAGATGCTGGGCATCCACTATCGGTTCGACGGTATCCAAGGCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACATCACTGTACGAACACTTCACCGAGGCTGATGAGCTTC 2123

RESULT 3
US-10-282-122A-41781
; Sequence 41781, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
```



```
Qy 84 LeuLeuProGlnAlaIleGlnValGlySerProThrHis---ProSerTyrProSerGly 102
|||:|||||
Db 268 CTGTCACCGCAGATCCGCCACCTCACC CGCGCAGCGCACACACCTCTCTTCCCGTCGGGA 327
|||:|||||

Qy 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuLysGlyLeuAspArg 122
|||:|||||
Db 328 CACTCGCCCTCGCGCGCGGCTTGGCAGCGC-----GTCGCCCTCGAGTCC 375
|||:|||||

Qy 123 GlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 142
|||:|||||
Db 376 ACCGGC-----

Qy 143 PheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuLeuLeuValAla 162
|||:|||||
Db 382 -----TACGGCGCACTGCTCGCGCGCTGCGCGCGCGCGGTGCGCC 420
|||:|||||

Qy 163 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeu 182
|||:|||||
Db 421 TTCTCTCGGCTCTATGTCGCGCTGCACTACCCCGGGGACGTCCTGCGCGGTATGCGCATC 480
|||:|||||

Qy 183 Gly 183
|||
Db 481 GGC 483
|||

RESULT 8
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 1.35e+04 Length: 9025608
Score: 93.00 Matches: 28
Percent Similarity: 39.60% Conservative: 12
Best Local Similarity: 27.72% Mismatches: 31
Query Match: 8.95% Indels: 30
DB: 15 Gaps: 3

US-10-691-383-2_COPY_435_632 (1-198) x US-10-156-761-1 (1-9025608)

Qy 84 LeuLeuProGlnAlaIleGlnValGlySerProThrHis---ProSerTyrProSerGly 102
|||:|||||
Db 8921985 CTGTCACCGCAGATCCGCCACCTCACC CGCGCAGCGCACACACCTCTCTTCCCGTCGGGA 8922044
|||:|||||

Qy 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuLysGlyLeuAspArg 122
|||:|||||
Db 8922045 CACTCGCCCTCGCGCGCGGCTTGGCAGCGC-----GTCGCCCTCGAGTCC 8922092
|||:|||||
```

```
Qy 123 GlyGlyGluCysPheProAsnProValPheProSerAspGlyLeuGluLeuLeuAsn 142
|||:|||||
Db 8922093 ACCGGC-----

Qy 143 PheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuLeuLeuValAla 162
|||:|||||
Db 8922099 -----TACGGCGCACTGCTGCGCGCTGCGCGCGCGGTGCGCC 8922137
|||:|||||

Qy 163 PheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeu 182
|||:|||||
Db 8922138 TTCTCTCGGCTCTATGTCGCGCTGCACTACCCCGGGGACGTCCTGCGCGGTATGCGCATC 8922197
|||:|||||

Qy 183 Gly 183
|||
Db 8922198 GGC 8922200
|||

RESULT 9
US-10-273-051-3
; Sequence 3, Application US/10273051
; Publication No. US20030157673A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Mario W.
; APPLICANT: Gibson, Katherine J.
; APPLICANT: Kostichka, Kristy N.
; APPLICANT: Thomas, Stuart M.
; APPLICANT: Nagarajan, Vasantha
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 US NA
; CURRENT APPLICATION NUMBER: US/10/273,051
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Rhodococcus ruber
; FEATURE:
US-10-273-051-3

Alignment Scores:
Pred. No.: 0.0297 Length: 1161
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservative: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: 15 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x US-10-273-051-3 (1-1161)

Qy 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
|||:|||||
Db 453 GCCCGCGGGTTTCGCGCGCGCTACCA CGCCAAACACAGCGCCACCTCTATCCGAGAAAT 512
|||:|||||

Qy 34 ArgPro-----GluAlaLeuGlyThrLeuHisAsnThr-IleAlaGlyAs 49
|||:|||||
Db 513 CGTCCCGCGCATCACC CGCGCGCGGTCAGGGAGTTCGTCCACACACATCCCGGTC 572
|||:|||||

Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAl 69
|||:|||||
Db 573 GCTCGCTCGCGACTTCAGATCGCGCGCGCGAGAACGAT-----TGGGCGCGTACCGC 626
|||:|||||

Qy 69 aGluLeuAsnAlaAlaGlnAsnProAsnGluValThrTyrLeuLeuProGlnAlaI 89
|||:|||||
Db 627 CGAGATCGTCGACCAACCCCTTCGATCGATCTCGCGCGCACTCGATCCGACACGCT 686
|||:|||||

Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
|||:|||||
Db 687 CATGGTAGACACGATGACGGGACCGGTAGCCGAGCGGACCGCGCCCATCTCCCGATG 746
|||:|||||
```


Qy 136 pGlyLeuGluLeuLeuLeuPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAen-- 155
Db 3338 CGGCTCGCTCTCTCTCACAGCACCTCGATCGGCAAGATTTTCGAAGAGCAGCAACAACGA 3397
Qy 156 ----LysLeuAlaValaAenValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172
Db 3398 CGTCGACCTCGCTCTCGGTGTCGCTTCGCGCGC-----GGCATCGGCTAC 3443

RESULT 12

US-10-129-518-13
; Sequence 13, Application US/10129518
; Publication No. US20030215930A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; TITLE OF INVENTION: Genes Involved in Cyclododecanone Degradation Pathway
; FILE REFERENCE: BC1023 PCT
; CURRENT APPLICATION NUMBER: US/10/129,518
; CURRENT FILING DATE: 2002-11-04
; PRIOR APPLICATION NUMBER: 60/170,214
; PRIOR FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 10480
; TYPE: DNA
; ORGANISM: Rhodococcus ruber
US-10-129-518-13

Alignment Scores:
Pred. No.: 0.821 Length: 10480
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservative: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: 16 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x US-10-129-518-13 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
Db 2951 GCCGGCGGTTCGGCCGCGCTACACCGCCCAACACACCGGCCACCTCATCGGAGAAAT 3010
Qy 34 ArgPro-----GluAlaLeuGlyGlyThrLeuHisAenThr-IleAlaGlyAs 49
Db 3011 CGTCGCGCGCATCACCGCCGCGACCGTTCAGAGAGTTCGTCCACACACATCGCGCGTCC 3070
Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgValAl 69
Db 3071 GCTCGCTCGGACTTCAGATCGCGCGCGCGAGAACGAT-----TGGGGCGCGTACCGC 3124
Qy 69 aGluLeuAsnAlaAlaGlnAenProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
Db 3125 CGAGATCGTCGACACCCCTTCGACATCGATCTCGCGCATCGATCCGACACCGT 3184
Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
Db 3185 CATGGTGAAGACGATACGCGGACCGGTAGCGACGCGAAGCGCGCAATATCTCCGATG 3244
Qy 102 -----GlyHisAlaThrGlnAenGlyAlaPheAlaThrValLeuLysAl 116
Db 3245 CGCGCACCGCACATGATGACCCCTCAACGGTCAACGCGCAACGCGCTCGCTCGCTCGAT 3304
Qy 116 aLeuLeuGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPheProSerAspAs 136
Db 3305 CCTTCACCATCATCGCTGGCGGTGAG-----ACCA 3337
Qy 136 pGlyLeuGluLeuLeuAenPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAen-- 155
Db 3338 CGGCTCGCTCTCTCCACAGCACCTCGATCGGCAAGATTTTCGAAGAGCAGCAACAACGA 3397
Qy 156 ----LysLeuAlaValaAenValAlaPheGlyArgGlnMetLeuGlyLeuHisTyr 172
Db 3398 CGTCGACCTCGCTCTCGGTGTCGCTTCGCGCGC-----GGCATCGGCTAC 3443

RESULT 13

US-10-425-114-22562
; Sequence 22562, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22562
; LENGTH: 1282
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3587-278-C11_FLI
US-10-425-114-22562

Alignment Scores:
Pred. No.: 0.0775 Length: 1282
Score: 89.00 Matches: 53
Percent Similarity: 34.86% Conservative: 23
Best Local Similarity: 24.31% Mismatches: 64
Query Match: 8.57% Indels: 78
DB: 13 Gaps: 11

US-10-691-383-2_COPY_435_632 (1-198) x US-10-425-114-22562 (1-1282)

Qy 11 LeuLeuGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29
Db 166 CTCACCGATCGCTTCTCGCGCGAGCGCGGGCGTGTTCCTTCGACCCATCCGGGTG 225
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAenThrIleAlaGly 48
Db 226 GTTCAACTCTCTTGTGGCGCCCGCGGGGTTC-----TTGCACAAGGGTTCTCTGTG 276
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgVal 68
Db 277 GATCGGAGTGCACACCATCTATCGCGCTGCCCAAGCAGCTGGAGAGTCTATGTTG 336
Qy 69 AlaGluLeuAsnAlaAlaGlnAenProAsnAsnGluVal----- 81
Db 337 GCGGACAACAAGTCCGGTAAGAGTGTCCAGAGTGAGGTGCGCACAGTTCGGCATGTT 396
Qy 82 -----Thr 82
Db 397 CTCGAGAAGAAGCAGGATGAAGTAGTAACGAGGATAGAGAGGATATCTGTTGTGACA 456
Qy 83 TyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySerPro 94
Db 457 TTTCTCCACAGAGATGTTGAGGCCATTCAGATATTACACTACCAAGATGGTGAGAA 516
Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAenGlyAla----- 109
Db 517 TATGAACCCCACTACGACTACTTCCATGACAAAAATAATCAGGCCCTCGGGCGGCATCG 576
Qy 110 PheAlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGlyGlu---CysPhePro 128
Db 577 ATTGCCCATGTGTGTATGTATCTATCAAAATGTTCAGAAAGGTGGAGAGACCATCTCCC 636
Qy 129 AsnPro-----ValPheProSerAspAsp----- 136
Db 637 AATGCAGAGGGGAAGCTATTTCGAACCCCAAGACACACTTGGTCTGATTGTGCAGAAAC 696
Qy 137 -----GlyLeuGluLeuIleAsnPhe----- 143

Db 697 GGATATGCGAGTAAACCGGTAAGGGTATGATCCCTGCTGCTTCTTCAGTCTCCACCCCTGAT 756
Qy 144 -----GluGlyAlaCysLeuThrTyrGluGlyGlu 153
Db 757 TCAACAACAGACTCTGACAGCTTGACAGCTAGTCCCGCTCATCGAAGGCCAG 810
RESULT 14
US-10-767-701-12438
; Sequence 12438, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 12438
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS9661_1
US-10-767-701-12438
Alignment Scores:
Pred. No.: 0.0988 Length: 1506
Score: 89.00 Matches: 54
Percent Similarity: 35.62% Conservative: 24
Best Local Similarity: 24.66% Mismatches: 61
Query Match: 8.57% Indels: 80
DB: 17 Gaps: 13
US-10-691-383-2_COPY_435_632 (1-198) x US-10-767-701-12438 (1-1506)
Qy 11 LeuGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29
Db 433 CTTCCGCGCTCGCGCTCTCGCGGAGCGCGGCGCTGTTCTTCGACCCCTCCCGCGTG 492
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGly 48
Db 493 GTCAACTCTCGTGGCGCCCGCCAGGCGCTTC-----CTGCACAAGGGTTTCCTGTCG 543
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLys---Arg 67
Db 544 GATCGGAGTGGAC---CACCTCATGCTGTGCCCAAGGACAACTGGAGAGTCTATG 600
Qy 68 ValAlaGluLeuAlaAlaGlnAsnProAsnAsnGluVal----- 81
Db 601 GTAGCGACAAACGAGTCCGGTAAGAGCGTCCAGAGCGAGGTGGCGCACCACTCCGGCATG 660
Qy 81 ----- 81
Db 661 TTCTCGAGAGAGAGAGAGATGATAGTAGTAGAGGGAATAGAGAGAGATAGTCTGTGG 720
Qy 82 ThrTyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySer 93
Db 721 ACATTCTTCCACAGAGAATGGTGAATCCATTACAGATATTACACTACCAGATGGTGAG 780
Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla----- 109
Db 781 AAATACGAACCCCACTACTACTCTCCATGACAAAAATAATCAAGCCCTGGGTGGCCAT 840
Qy 110 ---PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhe 127
Db 841 CGCATGCCACTGTGCTCATGTACTTATCTAATGTTGAGAGGGTGGAGACCATCTTC 900
Qy 128 ProAsnPro-----ValPheProSerAspAsp----- 136
::: ||| |||||

Db 901 CCAATGCAGAGGGAAGCTATTACAACCAAGGACACACATGGTCTGATTGTGCAAGA 960
Qy 137 -----GlyLeuGluLeuIleAsnPhe----- 143
Db 961 AATGGATATGCAGTTAAACCGGTAAGGGTATGATCCCTGCTGCTTCTTCAGTCTCCACCT 1020
Qy 144 -----GluGlyAlaCysLeuThrTyrGluGlyGlu 153
Db 1021 GATGCAACAACAGACTCTGAGAGCTTGCAGGTAAGTGGCCCCGTCATCGAAGGCCAG 1077
RESULT 15
US-10-169-395-14
; Sequence 14, Application US/10169395
; Publication No. US20040034192A1
; GENERAL INFORMATION:
; APPLICANT: KATO, Seishi
; APPLICANT: KIMURA, Tomoko
; TITLE OF INVENTION: HUMAN PROTEINS HAVING HYDROPHOBIC DOMAINS AND DNAS ENCODING
; FILE OF INVENTION: THESE PROTEINS
; FILE REFERENCE: 01997, 015100, US
; CURRENT APPLICATION NUMBER: US/10/169,395
; CURRENT FILING DATE: 2002-11-29
; PRIOR APPLICATION NUMBER: JP 2000-585
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: JP 2000-588
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: JP 2000-2299
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: JP 2000-26862
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: JP 2000-58367
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/JP00/09359
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 150
; SEQ ID NO 14
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-169-395-14
Alignment Scores:
Pred. No.: 0.171 Length: 1944
Score: 88.50 Matches: 38
Percent Similarity: 41.79% Conservative: 18
Best Local Similarity: 28.36% Mismatches: 39
Query Match: 8.52% Indels: 39
DB: 13 Gaps: 7
US-10-691-383-2_COPY_435_632 (1-198) x US-10-169-395-14 (1-1944)
Qy 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 280 GCGCGCTCAGTCCGCGCGCTCTGTATC-----CTCAGCAAC 318
Qy 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGly-----AspLeuAspAla 52
Db 319 CCCAG-----GGCACCTGTGAGGACACTCGGGCTGGCCCTCTCATGCA 369
Qy 53 AspPheAspIleSerLeuLeuGluAsnAspGlu-----Leu 64
Db 370 GACCACCTCTGCGCCCTGCTCGAGAGCCCAAGCCCTGACCCCGGCGCTGAGTGGCTG 429
Qy 65 LeuLysArgValAlaGluLeuAsnAlaAlaGlnAsnProAsnAsn-----GluVal 81
Db 430 CTGCAGAGATGACAGCCCGCGGCTCGCGGCCAGACCCCAAGACCGCGCTCGGTAGATATC 489
Qy 82 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSer 101
Db 490 CTTCACTGCTGAGAGAGCGGTGGGGGGGGGGCTCG----- 528
Qy 102 GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAsp 121
::: ||| |||||

Db 529 -----GGCAGTGTGTGGCGGCTCCTGTGCTGCCCTGTGACCATGTC 570
Qy 122 ArgGlyGlyGluCysPhe-----ProAsnProValPhe 132
Db 571 AGGAGCGGCTTGTCTTCCACGCCCTTGGCCGAGCCCTCAGTAC 612

Search completed: September 19, 2004, 11:18:46
Job time : 2394 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 04:00:38 ; Search time 42.5904 Seconds
(without alignments)
2579.931 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFLRIGABLAQR.....GLLGGTITVTLHQELMTF 198

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10691383/runat.17092004.102657.1766/app.query.fasta_1.1230
-DB=Issued_Patents_NA -QFMR=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10691383 @CIGN 1.1.93 @runat.17092004.102657.1766 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	2931	3	US-09-151-189-1
2	1039	100.0	2931	4	US-09-596-794-1
3	108	10.4	4403765	3	US-09-103-840A-2
4	108	10.4	411529	3	US-09-103-840A-1
5	99	9.5	51	3	US-09-151-189-3
6	99	9.5	51	4	US-09-596-794-3
7	91.5	8.8	1161	4	US-09-732-615-3
8	91.5	8.8	10480	4	US-09-732-615-13
9	87.5	8.4	1815	4	US-09-832-496-1
10	87.5	8.4	1815	4	US-09-832-616-1
11	87	8.4	1812	4	US-09-832-615-1
12	87	8.4	1812	4	US-09-832-617-1

13	85.5	8.2	1848	4	US-09-252-991A-11564	Sequence 11564, A	
14	84.5	8.1	1358	4	US-09-180-109A-25	Sequence 25, Appl	
15	83.5	8.0	1161	4	US-09-732-615-30	Sequence 30, Appl	
16	83.5	8.0	2822	2	US-08-679-405-1	Sequence 1, Appli	
17	83.5	8.0	2822	2	US-08-842-799-1	Sequence 1, Appli	
18	83.5	8.0	2822	5	PCT-US96-11458-1	Sequence 1, Appli	
19	82.5	7.9	873	4	US-09-134-001C-436	Sequence 436, App	
20	82.5	7.9	936	4	US-09-489-039A-5842	Sequence 5842, Ap	
21	82.5	7.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli	
22	81.5	7.8	1797	4	US-09-832-441-1	Sequence 1, Appli	
23	81.5	7.8	1797	4	US-09-833-102-1	Sequence 1, Appli	
24	80.5	7.7	1344	4	US-09-252-991A-16526	Sequence 16526, A	
25	80.5	7.7	2631	4	US-09-252-991A-16070	Sequence 16070, A	
c	25	79	7.6	26270	4	US-09-717-364A-1	Sequence 1, Appli
26	79	7.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl	
27	79	7.6	4411529	3	US-09-103-840A-1	Sequence 1, Appli	
28	79	7.6	726	4	US-09-252-991A-15963	Sequence 15963, A	
29	78.5	7.6	972	4	US-09-252-991A-16542	Sequence 16542, A	
c	30	78.5	7.6	1443	4	US-09-252-991A-16057	Sequence 16057, A
31	78.5	7.6	2094	4	US-09-489-039A-314	Sequence 314, Appl	
32	78.5	7.6	735	3	US-08-750-145A-23	Sequence 23, Appl	
33	76.5	7.4	735	3	US-08-975-698A-27	Sequence 27, Appl	
34	76.5	7.4	735	3	US-09-417-090-27	Sequence 27, Appl	
35	76.5	7.4	735	3	US-09-727-578-27	Sequence 27, Appl	
36	76.5	7.4	735	4	US-09-252-991A-11048	Sequence 11048, A	
37	75.5	7.3	1509	4	US-09-252-991A-11206	Sequence 11206, A	
c	38	75.5	7.3	1590	4	US-09-832-498-1	Sequence 1, Appli
39	75.5	7.3	1842	4	US-09-832-614A-1	Sequence 1, Appli	
40	75.5	7.3	1842	4	US-08-311-731A-137	Sequence 137, App	
41	75.5	7.3	40123	4	US-08-311-731A-137	Sequence 137, App	
42	75	7.2	924	3	US-08-952-089A-14	Sequence 14, Appl	
c	43	75	7.2	4146	3	US-08-952-089A-29	Sequence 29, Appl
44	74.5	7.2	1863	4	US-09-252-991A-3225	Sequence 3225, Ap	
c	45	74.5	7.2	2154	4	US-09-252-991A-3346	Sequence 3346, Ap

ALIGNMENTS

RESULT 1

US-09-151-189-1
; Sequence 1, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-09-151-189-1

Alignment Scores:			
Pred. No.:	4,94e-126	Length:	2931
Score:	1039.00	Matches:	198
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0
US-10-691-383-2_COPY_435_632 (1-198) x US-09-151-189-1 (1-2931)			

Qy 1 ValAspPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
|||||

Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGGCCGCGAGCTGGCGCAGCGT 1589
Qy 21 AlasSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCAGCGCCCGAGGGCTCTCGGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACAACACCATCGGGGGGATCTAGATGCAGACTTCGACATCTCCCTCTTTGAA 1709
Qy 61 AsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGTGGCGGAGATAAATCGGGCGAGATCCCAACACGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACCTACCTTCTTCACAAGCTATCCAAAGTGGATCGCAACGCGACCCCTTCTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGGAGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGGCTGGAACTA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGCGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCATTTGGAGGCGAGATGCTGGGCATCCACTATCGGTTCAGCGGTATCCAAAGGCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACAATCACTGTACGAACACTTCACCGAGGCTGATGACGTTTC 2123

RESULT 2

US-09-596-794-1
; Sequence 1, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2931
; TYPE: DNA
; ORGANISM: Fucus distichus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (228)..(2258)
; OTHER INFORMATION: vanadium bromoperoxidase
US-09-596-794-1

Alignment Scores:
Pred. No.: 4.94e-126 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x US-09-596-794-1 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGlnAsnArg 20
Db 1530 GTCAACTTCGGCAGCTCTCACTACTTCAGATTGATAGTGGCCGCGAGCTGGCGCAGCGT 1589
Qy 21 AlasSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
Db 1590 GCCTCGTGTACCAAAAGTGGCAGGTGCATCGATTTCAGCGCCCGAGGGCTCTCGGGGGT 1649
Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
Db 1650 ACCCTCCACAACACCATCGGGGGGATCTAGATGCAGACTTCGACATCTCCCTCTTTGAA 1709
Qy 61 AsnAspGluLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
Db 1710 AATGATGAGCTCTTGAACAGTGTGGCGGAGATAAATCGGGCGAGATCCCAACACGAG 1769
Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTCACCTACCTTCTTCACAAGCTATCCAAAGTGGATCGCAACGCGACCCCTTCTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGGCCACGCTACCCAAAATGGAGCATTTGCCACAGTTCTGAAGGGCCCTCATTTGGCCTA 1889
Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGGAGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGGCTGGAACTA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCAACTTCGAAGGGCGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCATTTGGAGGCGAGATGCTGGGCATCCACTATCGGTTCAGCGGTATCCAAAGGCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGGAGAGACAATCACTGTACGAACACTTCACCGAGGCTGATGACGTTTC 2123

RESULT 3

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 84.2 Length: 4403765
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72

```
DB: 3 Gaps: 8
US-10-691-383-2_COPY_435_632 (1-198) x US-09-103-840A-2 (1-4403765)
QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Dbb 4262534 GGACTCTCCCACTTCGGCGGAGCACTGATCGTTGGCTGATCCTGGCGCTCTCGGGCGG 4262475
QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHisArgPheAlaArg 34
Dbb 4262474 ATAGCCCTGCCACGGCGCCG-----CGGGAATGGCTTGTG-----4262439
QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Dbb 4262438 ---GCGGGGGCGGCGGCTTCGTCGCCCATGCCATCGCC-----4262403
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuIleValAlaGluIleAsnAlaAla 74
Dbb 4262402 -----GTGCTGATCAAGCGGCTGTGTGCGG-----CGT 4262376
QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Dbb 4262375 CAGCGGCGGATCATCGGCCATC-----GCGGTCAACGTGGACACGCCA 4262331
QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Dbb 4262330 AGTCAACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGGCTCTCATG 4262271
QY 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Dbb 4262270 GGCAGACCCACGGGCTG-----CGCTACCGGTTGTG-----4262238
QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Dbb 4262238 -----4262238
QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Dbb 4262237 -----CTGTGTCGGCGGATGCTGCGGAATACTGTCTGGGGGTCCACTACCCAGT 4262184
QY 175 AspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
Dbb 4262183 GATGTGCCGTGGGTGTGTCTGTGGCGCCACCGTC 4262148

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 84.5 Length: 4411529
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72

DB: 3 Gaps: 8
US-10-691-383-2_COPY_435_632 (1-198) x US-09-103-840A-1 (1-4411529)
QY 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Dbb 4270277 GGACTCTCCCACTTCGGCGGAGCACTGATCGTTGGCTGATCCTGGCGCTCTCGGGCGG 4270218
QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnIlystrpGlnValHisArgPheAlaArg 34
Dbb 4270217 ATAGCCCTGCCACGGCGCCG-----CGGGAATGGCTTGTG-----4270182
QY 35 ProGluAlaLeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Dbb 4270181 ---GCGGGGGCGGCGGCTTCGTCGCCCATGCCATCGCC-----4270146
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuIleValAlaGluIleAsnAlaAla 74
Dbb 4270145 -----GTGCTGATCAAGCGGCTGTGTGCGG-----CGT 4270119
QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Dbb 4270118 CAGCGGCGGATCATCGGCCATC-----GCGGTCAACGTGGACACGCCA 4270074
QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Dbb 4270073 AGTCAACTGAGCTTTCGTCGGCACACGCCACCTCGACACCGCGCGGCTCTCATG 4270014
QY 115 LysAlaLeuIleGlyLeuAspArgGlyGlyCysPheProAsnProValPheProSer 134
Dbb 4270013 GGCAGACCCACGGGCTG-----CGCTACCGGTTGTG-----4269981
QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Dbb 4269981 -----4269981
QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Dbb 4269980 -----CTGTGTCGGCGGATGCTGCGGAATACTGTCTGGGGGTCCACTACCCAGT 4269927
QY 175 AspGlyIleGlnGlyLeuLeuGlyGluThrIle 186
Dbb 4269926 GATGTGCCGTGGGTGTGTCTGTGGCGCCACCGTC 4269891

RESULT 5
US-09-151-189-3
; Sequence 3, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 51
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: probe for
; OTHER INFORMATION: second conserved region between Curvularia and
; OTHER INFORMATION: Ascopyllum vanadium peroxidase active sites
US-09-151-189-3

Alignment Scores:
Pred. No.: 2,79e-05 Length: 51
Score: 99.00 Matches: 17
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```


Alignment Scores:
Pred. No.: 1.05 Length: 10480
Score: 91.50 Matches: 50
Percent Similarity: 39.66% Conservative: 21
Best Local Similarity: 27.93% Mismatches: 74
Query Match: 8.81% Indels: 34
DB: 4 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x US-09-732-615-13 (1-10480)

Qy 14 AlaAlaGluLeuAlaGlnAraSerCysTyrGlnLysTrpGlnValHisArgPheAla 33
Db 2951 GCGGGGGTTCGGCGCGGCTACACAGCCCAACACAGCGCCACCTCATCGAGAAAT 3010

Qy 34 ArgPro-----GluAlaLeuGlyThrLeuHisAsnThr-IleAlaGlyAs 49
Db 3011 CGTCCGCGCATCACCGCGCGACGTTCAAGAGTTCGTACACACACATCGCGGTCC 3070

Qy 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAl 69
Db 3071 GCTCGTTCGCGACTTCAGATCGCGCGCGCGAGACGAT-----TGGGGCGGTACCGC 3124

Qy 69 aGluLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaI 89
Db 3125 CGAGATCGTCGACACCCCTTCGACATCGATCTCGCGCATCGATCCCGACAGCGT 3184

Qy 89 eGlnValGlySerProThrHisPro-----SerTyrProSer-- 101
Db 3185 CATGGTGAACAGCATGACGGGACGGTAGCGGCGGCAACGCGGCGCAATACTCCGGATG 3244

Qy 102 -----GlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 116
Db 3245 GCGGACGCGGACATGGGAGCCCTCAACGGTCAACGCGCAACGCGCTCGCTCGGAT 3304

Qy 116 aLeuLeuGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAs 136
Db 3305 CCTCTCGACCATCAGCTGGCGGTGAG-----AGCAA 3337

Qy 136 pGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluLeuAsn-- 155
Db 3338 CGGCTTCGGCTCTCTCACAGCACCTCGATCGGCGCAAGATTTTCGAGAGACGACACGA 3397

Qy 156 ----LysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyThrHisTyr 172
Db 3398 CGTCGACCTCGCTCGGTGTCGCGTTCGCGCGC-----GGCATCGGCTAC 3443

RESULT 9
US-09-832-496-1
; Sequence 1, Application US/09832496
; Patent No. 6503508
; GENERAL INFORMATION:
; APPLICANT: Danielssen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10042.200-US
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Dreschlera hartleibii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1815)
US-09-832-496-1

Alignment Scores:
Pred. No.: 0.231 Length: 1815
Score: 87.50 Matches: 32
Percent Similarity: 35.51% Conservative: 17
Best Local Similarity: 23.19% Mismatches: 38
Query Match: 8.42% Indels: 51
DB: 4 Gaps: 5

US-10-691-383-2_COPY_435_632 (1-198) x US-09-832-616-1 (1-1815)

Qy 71 IleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGln 90
Db 1126 CTCGGCGCGCGCCGACAAACACAAACAGATCCCATTC----- 1164

Qy 91 ValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110
Db 1165 -----AAGCGGCATTCCTCCCGCTACCCCTCTGGTCAACGGACCTTTGGCGGCGCGGTG 1218

Query Match: 8.42% Indels: 51
DB: 4 Gaps: 5

US-10-691-383-2_COPY_435_632 (1-198) x US-09-832-496-1 (1-1815)

Qy 71 IleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGln 90
Db 1126 CTCGGCGCGCGCCGACAAACACAAACAGATCCCATTC----- 1164

Qy 91 ValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110
Db 1165 -----AAGCGGCATTCCTCCCGCTACCCCTCTGGTCAACGGACCTTTGGCGGCGCGGTG 1218

Qy 110 ----- 110
Db 1219 TTCMAATGTGCGTAGATACTACAACGGCGCGTAGGAACATGGAACGACGACGAGCCA 1278

Qy 111 -----AlaThrValLeuLysAlaLeuLeuGlyLeuAspArgGlyGly 124
Db 1279 GACAACATTGCCATCGACATGGTAGTATCCGAGGAGCTAAACGGTCTAAGCCGTGACCTA 1338

Qy 125 GluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuLeu----- 141
Db 1339 CGCAACGCTACGATCCACCGCCCAATCGAAGACGAGCCCGGTATCGTACGACCCGC 1398

Qy 142 -----AsnPheGluGlyAlaCys--LeuThrTyrGluGlyGluLeuAsnLysLeu 157
Db 1399 GTCGTGCACACTTCAACTCGGCTGGGAACCTTATGTTCGAAAACGCAATTTACGAATT 1458

Qy 158 AlaValAsnValAlaPheGlyArgGlnMetLeuGlyThrHisTyrArgPheAsp 175
Db 1459 -----TTCCTGGGTGTCCACTGCGCGCTTTTCGAT 1485

RESULT 10
US-09-832-616-1
; Sequence 1, Application US/09832616
; Patent No. 6506586
; GENERAL INFORMATION:
; APPLICANT: Danielssen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10175.200-US
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: Dreschlera hartleibii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1815)
US-09-832-616-1

Alignment Scores:
Pred. No.: 0.231 Length: 1815
Score: 87.50 Matches: 32
Percent Similarity: 35.51% Conservative: 17
Best Local Similarity: 23.19% Mismatches: 38
Query Match: 8.42% Indels: 51
DB: 4 Gaps: 5

US-10-691-383-2_COPY_435_632 (1-198) x US-09-832-616-1 (1-1815)

Qy 71 IleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGln 90
Db 1126 CTCGGCGCGCGCCGACAAACACAAACAGATCCCATTC----- 1164

Qy 91 ValGlySerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPhe 110
Db 1165 -----AAGCGGCATTCCTCCCGCTACCCCTCTGGTCAACGGACCTTTGGCGGCGCGGTG 1218

Db 1252 TACTACAAACGGCGCGTGGCACTGGAAGAGAGCAACAGACCAACATCGCCATCGAC 1311
Qy 113 -----ValLeuLysAlaLeuLeuGlyLeuAspArgGlyGlyGluCysPheProAsnPro 130
Db 1312 ATGATGATATCCGAAGAGTTGAACGGCTCTCGCGGATTTGGCCCAACCTACACCCC 1371
Qy 131 ValPheProSerAspAspGlyLeuGluLeuLeu-----AsnPheGlu 144
Db 1372 ACGGCCCCCATCATCGGACCAACCGCGCTCGCGGACGGGTGTAGTCCGCCACTTCAGC 1431
Qy 145 GlyAlaCys---LeuThrTyrGluGlyGluLeuLeuLeuLeuLeuLeuLeuLeuLeu 163
Db 1432 TCTGCTGGGAATCATGTTTCAAAATGCAATCTCCCGCAAT----- 1473
Qy 164 GlyArgGlnMetLeuGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 175
Db 1474 -----TTTCTCGGTGTGCACTGGCGCTTTGAC 1500

RESULT 13
US-09-252-991A-11564
; Sequence 11564, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11564
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11564

Alignment Scores:
Pred. No.: 0.436 Length: 1848
Score: 85.50 Matches: 42
Percent Similarity: 32.98% Conservative: 20
Best Local Similarity: 22.34% Mismatches: 65
Query Match: 8.23% Indels: 61
DB: 4 Gaps: 7

US-10-691-383-2_COPY_435_632 (1-198) x US-09-252-991A-11564 (1-1848)
Qy 17 LeuAlaGlnArgAlaSerCysTyrGlnLysTyrGlnValHisArgPheAlaAArgProGlu 36
Db 1153 GTTGCCAGCGCGGAGGTGACGAGGATCAGCGC---CACGCGCTGCCAGCCCGAGAA 1209
Qy 37 Ala-----LeuGlyGlyThrLeuHisAsnThr 45
Db 1210 GGTGAAGCGCGCCAGGCTGCGGAGAACAGTCGACCTGGCAGGTACGCTGCACCGGTGA 1269
Qy 46 IleAlaClyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 65
Db 1270 GTAGCTGGTGGC----- 1281
Qy 66 LysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrLeuLeu 85
Db 1282 -----CAACGCGCGCAGCGCGCGAAGGAGAGATCACCAGGTGTGTATGACCGG 1332
Qy 86 ProGlnAlaIleGlnValGly----- 92
Db 1333 CCGCAGCGCGCGAAGCTGTGTCAGGCGAGGTTCGAGGTTCATCTGGGGCCAGACCATGTTG 1392
Qy 93 ---SerProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAla 111
Db 1393 CGAGCGGATCAGTACGAGCCCAACCCCATGCGCGATCAGCGCCCATACCGCGGTCTATGGC 1452

RESULT 14

US-09-180-109A-25
; Sequence 25, Application US/09180109A
; Patent No. 6410293
; GENERAL INFORMATION:
; APPLICANT: MUKUMOTO, Fujio
; APPLICANT: NISHIO, Shoichi
; APPLICANT: AKIMARU, Jiro
; APPLICANT: MITSUDA, Satoshi
; TITLE OF INVENTION: DNA Fragments Containing Biotin Biosynthesase Gene and
; FILE REFERENCE: 0152-0490P
; CURRENT APPLICATION NUMBER: US/09/180,109A
; CURRENT FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 09/047838 JAPAN
; PRIOR FILING DATE: 1997-03-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1358
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
; FEATURE:
; OTHER INFORMATION: Strain = SC42405
; NAME/KEY: CDS
; LOCATION: (152)..(1207)
US-09-180-109A-25

Alignment Scores:
Pred. No.: 0.366 Length: 1358
Score: 84.50 Matches: 57
Percent Similarity: 37.18% Conservative: 30
Best Local Similarity: 24.36% Mismatches: 80
Query Match: 8.13% Indels: 67
DB: 4 Gaps: 13

US-10-691-383-2_COPY_435_632 (1-198) x US-09-180-109A-25 (1-1358)

Qy 12 IleGlyAlaAlaGluLeuAlaGlnArgAlaSerCys---TyrGlnLysTyrGlnVal--- 29
Db 59 ATAAGGATGCGCAGCGCTGGATCGACGCGCGCTTGGCCCTATGACGGGGTTTGGCGCG 118
Qy 30 ---HisArgPheAlaArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAla--- 47
Db 119 CGTCATCCGCGCGCAGACGCGCGCTGAGGAATGACTATGACTACACCCCGCCATC 178
Qy 48 -----GlyAspLeuAspAlaAspPheAspIleSerLeuLeu 59
Db 179 ACTGCAGTACCGACTGGACCCCGTAGGAATTCGGCGCTGTTCGACCTGCCCGTTC--- 235
Qy 60 GluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsn 79
Db 236 -----ACCGAATGTTGTTCCGCGCAGCGCGAAGTCCATCGCGCCAGCCATCCGCAAC 289
Qy 80 GluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyr 99

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 03:58:58 ; Search time 2478.17 Seconds
(without alignments)
3463.005 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632
Perfect score: 1039
Sequence: 1 VNFGTSHYFLRIGAAELAQ.....GLLGLTITVRLHQLMTF 198

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10691383/runat.17092004.102656.1741/app.query.fasta_1.1230
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10691383 @CGN 1.1 6663 @runat.17092004.102656.1741 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.em.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1039	100.0	2931	6	AR152190	Sequence
2	1039	100.0	2931	6	BD235853	Gene-modi
3	1039	100.0	2931	6	AR435585	Sequence
4	1039	100.0	2931	8	AF053411	Fucus dis
5	503.5	48.5	3420	8	LDI491787	Laminaria
6	502.5	48.4	3379	8	LDI491786	Laminaria
7	332.5	32.0	1797	6	AX201842	Sequence
8	332.5	32.0	1908	8	AF218810	Corallina
9	331.5	31.9	2035	8	D87657	Corallina p
10	329.5	31.7	1794	6	E17199	Corallina p
11	322	31.0	1791	6	E17200	Corallina p
12	322	31.0	2029	8	D87658	Corallina p
13	116.5	11.2	12198	1	AE001953	Sequence
14	108	10.4	17205	1	AE007185	Myobacte
15	108	10.4	244800	1	BX842584	Myobacte
16	108	10.4	278492	1	BX248347	Myobacte
17	104	10.0	300450	1	AP005960	Bradyrhiz
18	101.5	9.8	11205	1	AE013888	Yersinia
19	101.5	9.8	199050	1	AJ414147	Yersinia
20	101	9.7	5828	6	E47157	Method for
21	99	9.5	51	6	AR152191	Sequence
22	99	9.5	51	6	BD235854	Gene-modi
23	99	9.5	51	6	AR435586	Sequence
24	99	9.5	302550	1	EX294137	Pirellula
25	97	9.3	302178	1	AE016918	Chromobac
26	93	9.0	325483	1	AP005050	Streptomy
27	92.5	8.9	1458	8	AY045606	Arabidops
28	91.5	8.8	1161	6	AR408895	Sequence
29	91.5	8.8	1161	6	AX167387	Sequence
30	91.5	8.8	10480	1	AY052630	Rhodococc
31	91.5	8.8	10480	6	AR408900	Sequence
32	91.5	8.8	10480	6	AX167397	Sequence
33	90.5	8.7	1389	8	AY139975	Arabidops
34	90.5	8.7	3107	5	BC054581	Danio rer
35	90.5	8.7	172289	2	AC146249	Pan trogl
36	89	8.6	37977	9	AC112718	Homo sapi
37	89	8.6	169828	9	AC097461	Homo sapi
38	88.5	8.5	1944	6	AX191492	Sequence
39	88.5	8.5	2117	6	AX083511	Sequence
40	88.5	8.5	2122	6	AX191502	Sequence
41	88.5	8.5	2178	9	BC062625	Homo sapi
42	88.5	8.5	2192	9	AK025537	Homo sapi
43	88.5	8.5	2248	9	AK000695	Homo sapi
44	88.5	8.5	3987	1	SAC550389	Sulfolobu
45	88.5	8.5	24985	1	AY372755	Bacteroid

ALIGNMENTS

RESULT 1

AR152190
LOCUS AR152190 2931 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6232457.
ACCESSION AR152190
VERSION AR152190.1 GI:15118240
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2931)
AUTHORS Vreeland, V. and Ng, K.L.
TITLE Recombinant vanadium haloperoxidases and their uses
JOURNAL Patent: US 6232457-A 1 15-MAY-2001;
FEATURES
source Location/Qualifiers
1. .2931
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores: 1.26e-92 Length: 2931
Pred. No.: 1039.00 Matches: 198
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6
US-10-691-383-2_COPY_435_632 (1-198) x AR152190 (1-2931)
QY 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
DB 1530 GTCAACTTCGGCAGCTCTCACTTTCAGATTGATAGTGCCTGGCGGAGCGT 1589
QY 21 AlaserCysTyrGlnIleValHisArgPheAlaArgProGluAlaLeuGlyGly 40
DB 1590 GCCTCGTGTACCAAAAGTGGCAGTGCATCGATTTCACGCCCGCGGAGTCTCGGGGT 1649
QY 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
DB 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTTCTGAA 1709
QY 61 AsnAspGluLeuLeuIleValHisArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
DB 1710 AATGATGAGCTCTTGAACCTGTGGCGAGATAAATCGCGGCGAGATCCCAACACGAG 1769
QY 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
DB 1770 GTCACCTACCTTCTTCCACAACTATCCAAAGTGGATCGCAACGACCCCTTCTACCCG 1829
QY 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIleGlyLeu 120
DB 1830 TCGGGCCAGCTACCCAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
QY 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
DB 1890 GATCGGGAGTGTAGTCTTCCCTAACCCGTGTTCCCAAGCATGACGCGCTGGAACCTA 1949
QY 141 IleAsnPheGluGlyAlaCysLeuThrTyrGlnIleValHisGlnLeuAlaValAsn 160
DB 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAGCTCGCGGTCAAC 2009
QY 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
DB 2010 GTCGCAATTTGGGAGGAGATGCTGGGCATCCATCATCGTTTCACGCGATATCCAGGCGCTA 2069
QY 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnLeuMetThrPhe 198
DB 2070 CTCTCGGAGAGACAATCACTGTACGAAACATTCACACGAGGCTGATGACGTTTC 2123
RESULT 2
BD235853 2931 bp DNA linear PAT 17-JUL-2003
LOCUS BD235853
DEFINITION Gene-modified vanadium haloperoxidase and utilization thereof.

BD235853
ACCESSION BD235853.1 GI:33045623
VERSION BD235853.1
KEYWORDS JP 2002525046-A/1.
SOURCE Fucus gardneri
ORGANISM Fucus gardneri
Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.
REFERENCE 1 (bases 1 to 2931)
AUTHORS Vreeland, V. and Ng, K.L.
TITLE Gene-modified vanadium haloperoxidase and utilization thereof
JOURNAL Patent: JP 2002525046-A 1 13-AUG-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT OS Fucus gardneri
PN JP 2002525046-A/1
PD 13-AUG-2002 JP 2000570298
PF 27-AUG-1999 JP 2000570298
PR 10-SEP-1998 US 09/151189
PI VALERIE VREELAND, KWAN L NG
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/02, C12N15/00, C12N5/00
CC Gene-modified vanadium haloperoxidase and utilization thereof
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1. .2931
/organism="Fucus gardneri"
/mol_type="genomic DNA"
/db_xref="taxon:3013"
ORIGIN
Alignment Scores: 1.26e-92 Length: 2931
Pred. No.: 1039.00 Matches: 198
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6
US-10-691-383-2_COPY_435_632 (1-198) x BD235853 (1-2931)
QY 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
DB 1530 GTCAACTTCGGCAGCTCTCACTTTCAGATTGATAGTGCCTGGCGGAGCGT 1589
QY 21 AlaserCysTyrGlnIleValHisArgPheAlaArgProGluAlaLeuGlyGly 40
DB 1590 GCCTCGTGTACCAAAAGTGGCAGTGCATCGATTTCACGCCCGCGGAGTCTCGGGGT 1649
QY 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
DB 1650 ACCCTCCACAAACACCATCGCGGGGATCTAGATGCAGATTCGACATCTCCCTTCTGAA 1709
QY 61 AsnAspGluLeuLeuIleValHisArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGlu 80
DB 1710 AATGATGAGCTCTTGAACCTGTGGCGAGATAAATCGCGGCGAGATCCCAACACGAG 1769
QY 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
DB 1770 GTCACCTACCTTCTTCCACAACTATCCAAAGTGGATCGCAACGACCCCTTCTACCCG 1829
QY 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIleGlyLeu 120
DB 1830 TCGGGCCAGCTACCCAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
QY 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
DB 1890 GATCGGGAGTGTAGTCTTCCCTAACCCGTGTTCCCAAGCATGACGCGCTGGAACCTA 1949
QY 141 IleAsnPheGluGlyAlaCysLeuThrTyrGlnIleValHisGlnLeuAlaValAsn 160
DB 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAGCTCGCGGTCAAC 2009
QY 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180

Db 2010 GTCCGATTTGGAGGACAGATGCTGGGCAATCCTATCGTTGCGAGGTATCCAAAGGCGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
|||||
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACACTTCCACGAGGCTGATGACGTTT 2123

RESULT 4
AF053411

LOCUS AF053411 2931 bp DNA linear PAT 18-DEC-2003

DEFINITION Sequence 1 from patent US 6656715.

ACCESSION AR435585

VERSION AR435585.1 GI:40198553

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2931)

AUTHORS Vreeland,V. Ng,K. and Epstein,L.

TITLE Recombinant minimal catalytic vanadium haloperoxidases and their uses

JOURNAL Patent: US 6656715-A 1 02-DEC-2003;

FEATURES
Location/Qualifiers
1..2931
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,26e-92 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AR435585 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
|||||
Db 1530 GTCAACTTCGCGACGCTCTCACTACTTCCAGATTGATAGTCCCGCGAGCTGGCGACGCT 1589

Qy 21 AlaserCysTyrGlnIleYsTrrGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
|||||
Db 1590 GCCTCGGTATTACCAAAAGTGGCAGTGTCATGATTTCCACGCCCGCGAGGCTCTCGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
|||||
Db 1650 ACCCTCCCAACACCATCGCGGGGATCTAGATGCACACTTCGACATCTCCTTCTTGAA 1709

Qy 61 AsnAspGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
|||||
Db 1710 AATGATGAGCTCTTGAACAGCTGTGGCGAGATAAATCGCGCGAGATAATCCCAACAGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
|||||
Db 1770 GTCACTTACCTTCTTCACCAAGCTATCCAAAGTGGGATCGCAACGACCCCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeu 120
|||||
Db 1830 TCGGGCCACGCTACCCAAATGGAGCAATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
|||||
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGCTGGAACCTA 1949

Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuLeuLeuLeuLeuLeuLeu 160
|||||
Db 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
|||||
Db 2010 GTCCGATTTGGAGGACAGATGCTGGGCAATCCTATCGTTGCGAGGTATCCAAAGGCGCTA 2069

Qy 181 LeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
|||||
Db 2070 CTCTCGGAGAGACAATCACTGTACGAACACTTCCACGAGGCTGATGACGTTT 2123

RESULT 4
AF053411

LOCUS AF053411 2931 bp mRNA linear PLN 12-SEP-1998

DEFINITION Fucus distichus vanadium bromoperoxidase mRNA, complete cds.

ACCESSION AF053411

VERSION AF053411.1 GI:3582762

KEYWORDS

SOURCE Fucus distichus

ORGANISM Fucus distichus

REFERENCE 1 (bases 1 to 2931)

AUTHORS Vreeland,V. Ng,K. and Epstein,L.

TITLE Direct Submission

JOURNAL Submitted (11-MAR-1998) Environmental Science, Policy and Management, University of California, 201 Wellman Hall, Berkeley, CA 94720-3112, USA

FEATURES
Location/Qualifiers
1..2931
/organism="Fucus distichus"
/mol_type="mRNA"
/db_xref="taxon:3012"
/dev_stage="2-cell embryo"
1..227
228..2258
/codon_start=1
/product="vanadium bromoperoxidase"
/protein_id="AAC35279.1"
/db_xref="GI:3582763"
/translations="MLCHAADTTTGGSPMDPTGVRLRLTSEQRAGMRRLQEGEKSLSGF
HPSETPVIKYLESETWKVKLTGDISASKILGKIMARVRIATATAVLAAPCLAFD
EVTASGVPEEHKHTGEGHLOCTNSDDALDPTAPNRDNVA PASRRDAARRERDGT
GTVCQTNGETDLATMFHKSILPHDELGCQTADDFAILEDCLNGDFSI CEDVPAGDPA
GRLVNTPAFAIDISGPAFSATTIPVPTLSSBELAAQLAELYNALARDVPPMQYGT
DEITTAANLANAGMGFPNLDVSI GSDGTVDPPFSQLFRATFVGVEGPPVSQLVNS
FTIDAITVPEPKQETFAFDNLNMYDFEMLNIQNGPPAGPEELDEELRFIRNARDLAR
VSPVDINTEAYRGSLLLELGAISRGPIIDSDROAGFVNFVSHYFRLTGAAB
LAORASYCKWQVHFARPEALGTLHTIAGDLDADFDI SILENDELLKRVAINAA
QNFNVTYLLPQAIQVGSPTSPSGHATQNGAFATVLKALLGLDRGCECFNPVF
PSDGLHLEFEGACITBEINKLVANVAFGRQMLGIHYRFDIGIQLLIGETITVTR
LHQLMTFAEEATFEFLFTGEVTKLFQDGTFTSIDGDMCSGLVYVTGVADCOA"

2259..2931

3'UTR

ORIGIN

Alignment Scores:
Pred. No.: 1,26e-92 Length: 2931
Score: 1039.00 Matches: 198
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-691-383-2_COPY_435_632 (1-198) x AF053411 (1-2931)

Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArg 20
|||||
Db 1530 GTCAACTTCGCGACGCTCTCACTACTTCCAGATTGATAGTCCCGCGAGCTGGCGACGCT 1589

Qy 21 AlaserCysTyrGlnIleYsTrrGlnValHisArgPheAlaArgProGluAlaLeuGlyGly 40
|||||
Db 1590 GCCTCGGTATTACCAAAAGTGGCAGTGTCATGATTTCCACGCCCGCGAGGCTCTCGGGGT 1649

Qy 41 ThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGlu 60
|||||
Db 1650 ACCCTCCCAACACCATCGCGGGGATCTAGATGCACACTTCGACATCTCCTTCTTGAA 1709

Qy 61 AsnAspGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
|||||
Db 1710 AATGATGAGCTCTTGAACAGCTGTGGCGAGATAAATCGCGCGAGATAATCCCAACAGAG 1769

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
|||||
Db 1770 GTCACTTACCTTCTTCACCAAGCTATCCAAAGTGGGATCGCAACGACCCCTTCCTACCCG 1829

Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLeuAlaLeuIleGlyLeu 120
|||||
Db 1830 TCGGGCCACGCTACCCAAATGGAGCAATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889

Qy 121 AspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
|||||
Db 1890 GATCGGGAGGTGAGTGCTTCCCTAACCCCGTGTCCCAAGCGATGACGCGCTGGAACCTA 1949

Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuLeuLeuLeuLeuLeuLeu 160
|||||
Db 1950 ATCAACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGTCAAC 2009

Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
|||||
Db 2010 GTCCGATTTGGAGGACAGATGCTGGGCAATCCTATCGTTGCGAGGTATCCCAACAGAG 1769


```

Qy 81 ValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 100
Db 1770 GTACCTTACCTTCTTCCACAGCTATCCAGTGGATGCGCAACGACCTTCTTACCCG 1829
Qy 101 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeu 120
Db 1830 TCCGCCACGCTACCCAAATGGAGCATTTGCCACAGTTCTGAAGGCCCTCATTTGGCCTA 1889
Qy 121 AsparGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 140
Db 1890 GATCGGGAGGTGAGTCTTCCCTAACCCCGTGTCCCAAGCGATGACGGCTTGGAACTA 1949
Qy 141 IleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsn 160
Db 1950 ATCACTTCGAAGGGGATGCTTACATATGAGGAGAGATCAACAAGCTCGCGGTCAAC 2009
Qy 161 ValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeu 180
Db 2010 GTCGCAITTCGGAGGAGATGCTGGGCATCCATCTATCGTTCGACGGTATCCAAAGGCCTA 2069
Qy 181 LeuLeuGlyGluThrIleThrValArgThrIleuHisGlnGluLeuMetThrPhe 198
Db 2070 CTTCTCGAGAGACAATCACTGTACGAACACTTCCACAGGAGCTGATGACGTTTC 2123

RESULT 5
LDI491787 3420 bp mRNA linear PLN 28-JUN-2003
LOCUS Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 2
DEFINITION (VBPO2 gene).
ACCESSION AJ491787
VERSION AJ491787.1 GI:32329413
KEYWORDS vanadium-dependent bromoperoxidase 2; VBPO2 gene.
SOURCE Laminaria digitata
ORGANISM Laminaria digitata
            Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
            Laminariaceae; Laminaria.
REFERENCE 1
AUTHORS Colin,C., Leblanc,C., Wagner,E., Delage,L., Leize-Wagner,E., Van
Dorselaer,A., Kiosreg,B. and Potin,P.
TITLE The brown algal kelp Laminaria digitata features distinct
bromoperoxidase and iodoperoxidase activities
J. Biol. Chem. 278 (26), 23545-23552 (2003)
JOURNAL 2703831
MEDLINE 22697758
PUBMED 2
REFERENCE 2 (bases 1 to 3420)
AUTHORS Leblanc,C.
TITLE Direct Submission
SUBMITTED (26-JUN-2002) Leblanc C., UMR 1931, CNRS-Goemar, Station
Biologique, BP 74, 29682 Roscoff, FRANCE
JOURNAL
FEATURES
            source
            1..3420
                /organism="Laminaria digitata"
                /mol_type="mRNA"
                /db_xref="taxon:80365"
                /tissue_type="sporophyte"
                /clone_lib="lambda ZAP II - S7"
                /country="France: Brittany"
            1..3420
                /gene="VBPO2"
            1..36
                /gene="VBPO2"
            37..2085
                /gene="VBPO2"
                /codon_start=1
                /product="vanadium-dependent bromoperoxidase 2"
                /protein_id="CAD37192.1"
                /db_xref="GI:32329414"
            /translation="MKEARPTQCALYSLRHHVACALCVACAFLLFEYQISSLGL
            TALLAFAPCLGVEEPEPTQPLLSGNVCVRVRSLSLDFDPVPRKVTLLKRLAIKDE
            ISVQPTCHVNNQDEENVLPFAGVHKTLPHDKFGQVDEDAIKLLCEVCSFDLINECK
            VPSGARRGAKLTNPJGGTAHQVTGADSDNVFTTDDSLLSRLAAQAEVYWMALL
            RDIPFGFAKNDYVRLAAENLQSLPAFKGLNI PRSEGGKIDPVDLFRITWPGVTTGP
            VWSQFMLSDFLIDSIKVTPKADPLTPGVDMYMTAFQPLWDVQNGASKLETTTDEENPRF

```

```

INRDLATIALRDOLYTEAFRAALILFTEGALGGEVGPVAEAEROQGCFATFGFPHIL
TAWASSTRHAWYAKQVHRMLRPEAYGALVHNTLRDVIITPLPDSILRNTELANR
VEVHNRMPDGEKIFLLPMAAQSPTPAIPSGHAINNGAYITLAKAFLGYEAGQK
CFNPVSVNDEGKIRIKYPSREIVGECNKGKLVGGLITYEGSLNKISANVLLGRS
HIGVHRMDGVYALMGKETS CVRRLOQLPGLPEAREVEGKKRRGDI PPATYKFRLYS
GKILELYGNLYKLDGLKCEGATGDDFCDDPIDDEDYSSFFDIVEHAOFSLHHTLEL
"
2086..3420
/gene="VBPO2"

3'UTR
ORIGIN
Alignment Scores:
Pred. No.: 1,01e-39 Length: 3420
Score: 503.50 Matches: 98
Percent Similarity: 60.00% Conservative: 28
Best Local Similarity: 46.67% Mismatches: 67
Query Match: 48.46% Indels: 17
DB: 8 Gaps: 2

US-10-691-383-2_COPY_435_632 (1-198) x LDI491787 (1-3420)
Qy 3 PheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSer 22
Db 1192 TTCCGGGAGCGGCACATCTCTCACCGCATGCGCTCGGCCAGCTCTTCCACGCGCCACGCG 1251
Qy 23 CysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeu 42
Db 1252 TGTAGCTGTAAGTGGCAAGTCCACCGCATGCTCGACCGCCGAGGCGTACGCTGCGCTGGTA 1311
Qy 43 HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAsp 62
Db 1312 CATAACAGCTCATGAGGAGCGTCATCCCCGCTCGCGGACTCTATCTCTGAGGAACACG 1371
Qy 63 GluLeuLeuLysArgValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThr 82
Db 1372 GAGCTCTCTCAACCGAGTGGAGGTGCACACCGCATGATGAACCCAGACGCGCGAGAAGACC 1431
Qy 83 TyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGly 102
Db 1432 TTCTGTCTCCCTATGCGGAGCGCCAGGCTCTTCCACGACCCCGCTTACCCAGTGGGA 1491
Qy 103 HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArg 122
Db 1492 CACGCCATCAACAACGCGCTACATCACCGCTCAAGCGTTCTCTCGGTACGAGGCC 1551
Qy 123 GlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeuIleAsn 142
Db 1552 GGCCAGAAGTGTCTCCCAACCCGCTGCTCGAACGACGAGGAACCAAGAGGATCAAA 1611
Qy 143 PheGlu-----GlyAlaCys----- 147
Db 1612 TACAGCCCTCCGAGGGAGATTGTGGGCGAGTGTCAACGAGAGGCGCAAGCTCGTC 1671
Qy 148 -----LeuThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArg 165
Db 1672 GAAGGCTTCACGTACGAGGAGAGAGTCAACAAGATCAGCGCCCAACGCTCTTTGGGAAGG 1731
Qy 166 GlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 185
Db 1732 TCCCATCATCGTGTCTCACTGGAGGATGGACGGCTGTACGAGGCGCTCATGGGAGAGACC 1791
Qy 186 IleThrValArgThrLeuHisGlnGluLeu 195
Db 1792 AGCTGGTCCGCCCTGCGAGGAACTT 1821

RESULT 6
LDI491786 3379 bp mRNA linear PLN 28-JUN-2003
LOCUS Laminaria digitata mRNA for vanadium-dependent bromoperoxidase 1
DEFINITION (VBPO1 gene).
ACCESSION AJ491786
VERSION AJ491786.1 GI:32329411
KEYWORDS vanadium-dependent bromoperoxidase 1; VBPO1 gene.

```


AUTHORS Shimonishi, M., Kuwamoto, S., Inoue, H., Wever, R., Ohshiro, T., Izumi, Y. and Tanabe, T.
TITLE Cloning and expression of the gene for a vanadium-dependent bromoperoxidase from a marine macro-alga, *Corallina pilulifera*
JOURNAL FEBS Lett. 428 (1-2), 105-110 (1998)
MEDLINE 98307393
PUBLISHED 9845486
REFERENCE 2 (bases 1 to 2029)
AUTHORS Tanabe, T.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1996) Tadaashi Tanabe, National Cardiovascular Center Research Institute, Department of Pharmacology; Fujishiro-dai 5-7-1, Suita, Osaka 565, Japan
FEATURES (E-mail: tanabe@ri.ncvc.go.jp, Tel: 06-833-5012, Fax: 06-872-8092)
 Location/Qualifiers
 1..2029
 /organism="Corallina pilulifera"
 /mol_type="mRNA"
 /db_xref="taxon:78447"
 /clone="BPO2"
 119..1912
 /codon_start=1
 /product="vanadium-dependent bromoperoxidase 2"
 /protein_id="BAA31262.1"
 /db_xref="GI:3293052"
 /translation="MGIPADNLOSRAKASFDTRVAAAEALALARGVPSFANGBELLYR
 NCTGDSFTASFTKGLPHDNGAIIIDPDLFLAFVRAINSDEKEIADLTGPARDPE
 TGLPIWRSDIANSLELRVGNWSSAGLTFLDGPDAQSVMPAPVLMSPELIAEVA
 ELYMLGRDIESEFEPSPKNAAFIRSAIERLNGLEWNTPAKLGPPAIEIRRGVH
 TVGNLFRGILPSEGVLYSOYIIVGSKQIGSATVGNKTVFVSPADEFGEIAYGSI
 TISQVRVIAITGDFMDLKVLDVQDGFGRFSEYEPGARLIRITDIATVWHFDA
 LYEAYLNACILILANGVPFDPNLPFOQEDKLDNDHVFVNFSGARHLSLVEVAWRAK
 AVRYKFIHVRRLPEATGGLISVNRKSLAGSDIIIPVEVSELVELSSILFDVAESN
 EKONRADIIVSRPKSFLPLMAFAEGSPFSPVSGHVAVAGACVTLILKAFDFANFQID
 KVEYVDDEKLVKSSFKGLITVAGELNKLADNVAIGRNAGVHYFSDQFESILLGEQ
 IAIGLIEQSLTYGTFNFFNLKPDGTTIIQI"
ORIGIN
 Alignment Scores:
 Pred. No.: 4,38e-22 Length: 2029
 Score: 322.00 Matches: 87
 Percent Similarity: 52.97% Conservative: 29
 Best Local Similarity: 39.73% Mismatches: 71
 Query Match: 30.99% Indels: 32
 DB: 8 Gaps: 8
 US-10-691-383-2_COPY_435_632 (1-198) x D87658 (1-2029)
 Qy 1 ValAsnPheGlyThrSerHisTyrPheArgLeuIle---GlyAlaAlaGluLeuAlaGln 19
 Db 1232 GTGAACCTCGGATCCGCACACGCTGCTGAGTCTGCTGACTGAAGTGGCTACGCGCGCGTTG 1291
 Qy 20 ArgAlaSerCysTyrGlnIlystrPglNValHisArgPheAlaArgProGluAlaLeuGly 39
 Db 1292 AAGCGGTACGGTATCAAGAGTTTAAACATTCATCGCTCGCTCGCGCTGAGCGTACTGGT 1351
 Qy 40 GlyThrLeu-----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAsp 55
 Db 1352 GGCTGATTAGCTTAAACAGAAATCTTCTTCGCGGTAGT-----GAT 1396
 Qy 56 IleSerLeuLeuGluAsnAspGluLeuLeuLysArg-----Val 68
 Db 1397 ATAATATTCCTGAAGTTAGTGAACCTGGTCGAGGAGCTCTCATCAATTCGTGATGACGTT 1456
 Qy 69 AlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal-----ThrTyr 83
 Db 1457 GCTGAGAGCAATGAAACACAGACAGCGGCTGACGGGATTGTAGCCCGGATAAATCATTT 1516
 Qy 84 LeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHis 103
 Db 1517 CTGTTGCGGATGGCATTTTGGGAGGAGGAGCCCATTCATTCGTCCTCTATGAGAGTGGCCAC 1576
 Qy 104 AlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly----- 119

Db 1577 GCTGTGGTTCGTCGCGCATGTGTGACAATCTCAAGCGGTTCTTCGACGCCAATCTCCAG 1636
 Qy 120 LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAspAspGlyLeuGlu 139
 Db 1637 ATCGATGAAGTGTTCGAG-----GTCACACTGATGAGGACAACTGTGTG 1681
 Qy 140 LeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaVal 159
 Db 1682 AAGTCGCTTTTCAAGGGAACCT---CTCACTGTTGCCGTTGAATGAACAAGCTCGCCGAC 1738
 Qy 160 AsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPheAspGlyIleGlnGly 179
 Db 1739 AATGTTGCGATCGCGCGGAACATGCGCGGTGTTCTCACTACTTCTCTGACCACTTCGAGTCA 1798
 Qy 180 LeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe 198
 Db 1799 CTGCTACTCGGTGAACAGATTCGATTCGATTCGATTCGAGGAGCAGACCTGACGAT 1855
RESULT 13
LOCUS AE001953/c
DEFINITION Deinococcus radiodurans R1 section 90 of 229 of the complete chromosome 1.
ACCESSION AE001953 AE000513
VERSION AE001953.1 GI:6458740
KEYWORDS
SOURCE Deinococcus radiodurans R1
ORGANISM Deinococcus radiodurans R1
REFERENCE 1 (bases 1 to 12198)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
TITLE Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 2036896
PUBLISHED 10567266
REFERENCE 2 (bases 1 to 12198)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES Location/Qualifiers
 1..12198
 /organism="Deinococcus radiodurans R1"
 /mol_type="genomic DNA"
 /strain="R1"
 /db_xref="taxon:243230"
 /chromosome="1"
 92..997
 /gene="DRI1011"
 /gene="DRI1011"
 /note="similar to GB:AE000657 percent identity: 62.08; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAFI0587.1"
 /db_xref="GI:6458741"
 /translation="MTSSAPESALPDVLHLGVSFCPNDFIFIALHAGLVQGLPVG
 EVLEVTQLNDNAVAGLPMTKISYRAYFDVMDRYVALRSGGALRGVGLIVTRPGV


```

KDLTALRVASPGALTTAELLRLAYPGVEVVRMYRDEVPVAVQGEYDGTAVDAGLI1
HESRFTYPEYGLDKLLDGLAWGEGTGLPLPLGAILVRDLPAQTORELOQAVRRSL6
YAYAPAAADYIREHALEMSDEWQAHIDLYNLSLDVGESEGERAVRELLRAVEV
GAVTARELPFLVGDAAEAGTPTIL"
1033..1956
/ gene
CDS
/ gene="DR1012"
1033..1956
/ gene="DR1012"
/ notes="similar to GB:AL009126 percent identity: 54.74;
identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="ABC transporter, ATP-binding protein"
/ protein_id="AAFI0588.1"
/ db_xref="GI:6458742"
/ translation="MSKASAPAPPAVEVRGLSKSYGRANVLTDFLVNVPGEVYALTG
PNAGKTTLRIMTGLAFTPRGEVRGNVNVHTDGRRAALMGAVAPAKFYQFTG
TENLKHASITAGVGGRISEVLALLELTSMGGRVABEYSLGQRLRGVASAILAEP
KVLIDPRTSGLDPLGTLIRHIVTSLATSGCAVILSTHLREIATVAHVTGIMTGR
LVTDVLRSTSVARFRVDDPVGAALLERQPYVRKATPRPPYAVAQLGAQANVPEAL
RLANEDVAYEYTPDHFDLXEYRVEYA"
1956..2729
/ gene="DR1013"
1956..2729
/ gene="DR1013"
/ notes="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAFI0594.1"
/ db_xref="GI:6458748"
/ translation="MLSLIERKLLASRSARIALVCTLLPLPMSFAPRLGEILRV
DIVSGQLPAVSGVAVQFLMPLFIAVAEMIGAEVAQCTLAPLLLRVRSRVRIVS
KLIVALFPULLLVALLGLSLVAGIPHYGEFRGGTGLGFLQGVGVLSFSAALAEV
LRGLGLTWIMLPMVAALSLFYGLLSTSAALATLATLSVLRLLVFPPELQORVLLT
SYNLXYTQSLTQALVLLIYTLGFSLMAIFAPDRDL"
complement(2935..3372)
/ gene="DR1014"
complement(2935..3372)
/ gene="DR1014"
/ notes="similar to GP:188406 percent identity: 57.35;
identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="vanadium chloroperoxidase-related protein"
/ protein_id="AAFI0589.1"
/ db_xref="GI:6458743"
/ translation="MTPGLMIIEALRGELQARLGSDLAQVLAATAVAGHDAFISCW
QKGFYNNVAPQSMMDHVQPCWAPSLPTPPFPSPYSGHATVSGAAAEVLAQFPPLQAR
QLRRDADAASRVVGGIHNGVDGAGLDVGQVRVARALLEKP"
complement(3423..4277)
/ gene="DR1015"
complement(3423..4277)
/ gene="DR1015"
/ notes="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAFI0595.1"
/ db_xref="GI:6458749"
/ translation="MSRASLSHWQPHRRRLARWCASVCLGIAAGAGAAQSSAVQSL
TAPSLPQTGAPALSPAVRSGLAASPLVISPAPXLRQTRRSALAAAPAGPATALSVO
VPRVARITDLDLTAARLDPPPARALATLGLAHDAWLARQSGRAVDSDTAAVNA
TRVQALFQKDALSELKSALFPLPAEVDRAASGLAVADAVLAPAQDQAQNAS
QGNAAQPSGSLWAPLSGQRALEFGWGRVQPIGLAPDALPQSRPPRAGRAPSPPTAS
VSPTSREP"
4376..5707
/ gene="DR1016"
4376..5707
/ gene="DR1016"
/ notes="similar to SP:P25755 PID:45710 percent identity:
57.11; identified by sequence similarity; putative"
/ codon_start=1

```

```

/transl_table=11
/ product="thiophene and furan oxidation protein"
/ protein_id="AAFI0590.1"
/ db_xref="GI:6458744"
/ translation="MTFARSSRSGLOTTAAVATAGSAGVGIVRVSGPRALBIADGL
FRKRRSATPGGRFLFGFHADDELLDEGLCLVFRGPRSYTGEDVABVTHGSPAVL
SRVLAITDLGARLARPGFTTRAYLAGRLDITQAEAVLGEVASTETARQSALGLS
GALSERVAGVARALTRTAAIQALDYPEEGVDEDRAPPLGAQSALELLASAHAG
RSTRCARLALIGRNAGKSSLNALLGVRSIVTPLGCTTRDYLEAGLELAGVPVTL
VTAGLRETDLEIAAGQVQAVNALAGNADLVLVLEDSQSPREPLTPELTPTATVLR
TKADLPAAWTDPAALDVSAVTGAGLPALREAIGALLGDATRGEMLTTRQDITVRR
ALTHVQAARTLPDDLAGELEELHALAEELTQGDQVEDVVDAVFRNFCVGK"
6131..6928
/ gene="DR1017"
6131..6928
/ gene="DR1017"
/ notes="similar to PID:1653441 percent identity: 51.76;
identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="rRNA methylase"
/ protein_id="AAFI0591.1"
/ db_xref="GI:6458745"
/ translation="MPDSLPHVITSQNQVQKLVRLTRRRERDGMIEGARELM
RAESAGVTPPELFCPELSPARELPFVETGATELSRAAFKVSRENPDGVLGVAP
LPAPRLPQGGDITVVVVLHGLEKPGNVGAILRTADAAGAGVLVLRGADPGYGNVIR
ASQAVSVPTAVLGEDAALGWLAEHNFRTVACTPDAPADYWDTPLTGRVALLGTEH
AGLPAEWRAAHQAVRVMGQADSLNVATAALVLYEALRQRRPAD"
7131..8327
/ gene="DR1018"
7131..8327
/ gene="DR1018"
/ notes="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAFI0596.1"
/ db_xref="GI:6458750"
/ translation="MSAPSPASQDPYRDLRLGLRGFGVREADOLIOAAVORRGWQ
AVRVLGPRDVAVLVQDYQMRDENGENDRADWLQASVDLNLAEVTPVPVDPEDV
LURPAPRPVANGRRADLPMLARVHEIATRSLSIRTDADLAPLERLRAAEWDVQ
ATQAEVRWETEDMLAQRLSQHARTVTDQVASARQVELLELTRELAAALTGERSSP
AKLAHNLMTQTRSFLEDAFVPLVNEQADQPLTMDIDLRHARYSLGVPLHPNVLRA
RYGLNYALWQAGELGRVAVQEAQALAQAEAAQLRTTTEAARTHQSTLQALAA
HYELEERALKLSRVGDDLSLARVLRLEWRQARAAARIQANRLEEAVALLEALVSDSA
EYSTLG"
8492..10060
/ gene="DR1019"
8492..10060
/ gene="DR1019"
/ notes="similar to GB:AE000783 percent identity: 68.44;
identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="glycerol-3-phosphate dehydrogenase"
/ protein_id="AAFI0592.1"
/ db_xref="GI:6458746"
/ translation="MTQGSVSDPRPAQVRAALAPQEWVDVIVVGGSGSLGSAVERAATR
GHSVLLLEGHDYAKGTSSTKLVHGVRVYLAQGNVSLREALHERGLLRKNAPHLVR
DLGVFPYQDWAAAPFYGLKMYDLAGLNIGSSTYLDRLDKLVRSVSTYKKGKLRG
GILYFPQDQDLARLALTLLRTLEDFGVALNYAPVGLIKDGEQVGRVFRDESGQE
HEARAKIVVNATGVWVDITRRMEDPTVKPMSPSQGVHVVKKFLPGDSAIMIPTID
DGRVLPAVPHDHVIGTDTTPVQADFEFHLEEEVEFILTAGRMYDPAGRADVL
SVYAGRLPLVNEKSGAGSTAALSRDHVRIRISKNLITLTGCKWTTWRMGSDTINR
AEKLGELPRLTITPGLHLHGWSEDERPDHWKVIGNDAERVQQLPGADAGHDLPLFS
EAELRWAVRHESATVEDVLARTRISILLGARASMEAPRAAAILAEELGKDAAWQAO
QVEAYRQVAGSYVL"
complement(10120..12087)
/ gene="DR1020"
complement(10120..12087)
/ gene="DR1020"
/ notes="similar to PID:1652085 percent identity: 67.48;
identified by sequence similarity; putative"

```


AIAGGVATKIPLSKWFLIMAFGSLFMVAGKRYVAELHIAERTGAARIKRSLESYSTVL
 RFVWTSATAVLCYGLWAFERDYGSGWFAVSMIPTIALLRYAVDVGGLAGEPED
 IALRDVQLQALLALAIATVCAAVAFG"
 /gene="MT3914"
 CDS
 complement(5407. .5904)
 /gene="MT3914"
 complement(5407. .5904)
 /gene="MT3914"
 /note="identified by match to PFAM protein family HMM
 pf01569"
 /codon_start=1
 /transl_table=11
 /product="PAP2 superfamily protein"
 /protein_id="AAK48280.1"
 /db_xref="GI:13883798"
 /translation="MVAVQSALVDRPGMLATARGLSHFGEHCIGWLILALLGALALPR
 RREWLAVAGAGFAVAHAIALLRLVRRQRPDHPATAVNDVTPSQLSFPPSAHAISTIA
 AALLMGSRATGLPVLVLPVPMALSRILLGVHYPDSVAUGVALGATVGAIVDSVGGGRQ
 RARRR"
 complement(5933. .7846)
 /gene="MT3915"
 CDS
 complement(5933. .7846)
 /gene="MT3915"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48281.1"
 /db_xref="GI:13883799"
 /translation="MSLAAALLSRVLPRGPBLVVRKLYLEESTNNARRAHAPTRT
 SLQIGAESEVSFATYFNAFPASVYRWTTCKSVLVVQVTVAGRVVYRTKATGARIF
 VEGHDFGTGDEPAAEVTEVLQPFEDGGMVFDITDTATVILHSGWATSPAPGTA
 NIAGVIFTEVRPADCVNARELTALVDQVIGAVIVPDQGERKVRDHPDFPAAARL
 GSRLSHDQNLGSGGYSRVMEALKNTCCQILPMDDIRLEPDSILLVLAHREPA
 KAPLVGGMLNQLQEPHLMGEVSDRSIFMTTAPHAEDYDHFAPYPLNDNRSRK
 LHRHIDVDYNGWTCMPROVABELGQPLFIKWDADYGLRAEHEGYPTVTLPGA
 AIWHMAWSDKDDAIDWQAYFHLNRNLVVAAMHWDPKAQVIGLVRSLKATLKLACL
 EYSTVAIQKAIIDDFLAGPEHIFSILESALPRVHRKSPYDAVLPAASELPPHLK
 NKAKPPVNLVIGYRLARGIMENLTAAHPQHRREFNVPTODARWFLICTVDGATV
 TTADGGVVYRQDRAKMFWALLWQSLRQRQLLKRFEEMRIYRDALPTLSSKQWET
 ALLPAAHQEPFHG"
 complement(7843. .9042)
 /gene="MT3916"
 CDS
 complement(7843. .9042)
 /gene="MT3916"
 /note="similar to SP:P37747 GB:U03041 GB:U09876 PID:508242
 PID:510253; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="UDP-galactopyranose mutase"
 /protein_id="AAK48282.1"
 /db_xref="GI:13883800"
 /translation="MQPMTARFDLVVVGSGFFGLTIAERVATQIDKRVLVLRPHIG
 GNAYSEAPGTGIEVHKYGAHLTSNKRVDYVQRPDTDFYHRVFMHNGQAVQF
 PMGLGVSPQFGFYFTQEQARLTAQAEIADTAQNLBEKAIISLIGRLPYEAFVKQ
 YTAQWOTDKPELPAAMTLPVYTFDNRYSFTYEGPLDGTGVTQNLQNAADHRIE
 VRLTNDFFDVGOLRCSRAAPVYVGTGLDPLDYAEGRLGWTLPDEVEVLPIGDFQ
 GTAMVYNDLDVYTRIHREHPEHREDYTDKTVIMREYSRAEDDDPEYPIINTEA
 DRALLATYARAKSETASSKVLFFGRIGTYIYLDMEALIASALNMTNDVLAHLRDGV
 PLIQDGA"
 /gene="MT3917"
 CDS
 /gene="MT3917"
 /note="similar to PID:845583 GB:AL123456; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="exported repetitive protein"
 /protein_id="AAK48283.1"
 /db_xref="GI:13883801"
 /translation="MPNRRKKUSTAMSAVALAVASPCAYFLVYESTETTERPEHHE
 FKQAALVDLPGLMSALSQLGQFGINIPVPSLTGSGDASTGLTGPGLTSPGLTSP
 GLTSPGLTDPALTSPGLTPTLPGLAAPGTTLTAPTCGVGANPALTNPALTSPTGATPG

LTPSPTGLDPALGANIEPITTPVGLDPGADGTYPIILGDDPTLTGTPSPSPATTSTGGGL
 VNDVMVANELGASQAIDLLKGVLMPSIMQAVQNGGAAAPAAAGPPVPIPAAGAAVPT
 DPTTPVA"
 /gene="MT3918"
 CDS
 /gene="MT3918"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48284.1"
 /db_xref="GI:13883802"
 /translation="MAATVIVAMIANRRPASSHEPSTENTOLASOPLILGCGVTV
 RELTQTPSLVALTGLDLAGTSARVAKRPDGDWGPWYOTETEYERDPAQTPGSSVEL
 GGNLPRSDPDPVFGVTTTVQVAVTRFDAPITQDPAGRPNDLDSGLGYPATYKE
 QPFGONISAILISPPQAPGTQMTPTATVMAGPPPIISRBAWGADESURCTEPEYD
 RGVRAAVVHTHAGSDNYSPLSAGIVKAIYTHSKTLGWCMDIAYNALVDKYGQVFFGS
 AGGLTPVGEFHTGGFNRTWGMIGNFDDVAPTPIQIRTVGRLLGWRILGMDVDPR
 SMVDLSAGSSYTFPGGAIARLPAIFTHRDVGNTPCPGNAAYVAMDEIRDIAAHFND
 Alignment Scores:
 Pred. No.: 7.1 Length: 17205
 Score: 108.00 Matches: 46
 Percent Similarity: 35.42% Conservative: 22
 Best Local Similarity: 23.96% Mismatches: 52
 Query Match: 10.39% Indels: 72
 DB: 1 Gaps: 8
 US-10-691-383-2_COPY_435_632 (1-198) x AB007185 (1-17205)
 QY 4 GlyTh-SerHisTyr-----PheArgLeuLeuGlyAla 14
 DB 5847 GGACTCTCCCACTTCGGCGAGCACTGCATCGGTGGTGATCTCTGGCGCTGCTGGGCGG 5788
 QY 15 AlaGluLeuAlaGlnArgAlaSerCysTyrGlnLysTyrGlnValHisArgPheAlaArg 34
 DB 5787 ATAGCCCTGCCACGGCGCGC-----CGGAATGGCTGTGTG----- 5752
 QY 35 ProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
 DB 5751 ---GCGGGGCGGGCGGCTTCGTCGCCCATCGCATCGCC----- 5716
 QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 74
 DB 5715 -----GTGCTGATCAGCGGCTGGTGGCG-----CGT 5689
 QY 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
 DB 5688 CAGCGGCGGATCATCGGCCATC-----GCGGTCAACGTGGACACGCCA 5644
 QY 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
 DB 5643 AGTCAACTGAGCTTCCTCGGCACACGCCACCTCGACACCGCGCGGCTGCTCATG 5584
 QY 115 LysAlaLeuIleGlyLeuAspArgGlyGlyGluCysPheProAsnProValPheProSer 134
 DB 5583 GGCAGAGCCACCGGGCTG-----CCGCTACCGGTGTGTG----- 5551
 QY 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
 DB 5551 ----- 5551
 QY 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
 DB 5550 -----CTGTGCGCGCGATGCTCGCGAATACTGCTGGGGGTCCACTACCCCACT 5497
 QY 175 AspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 186
 DB 5496 GATGTGCGCGGTGGTGTGTGCTCTGGGGCGCCACCGTC 5461
 RESULT 15
 BX842584/c

```

LOCUS      BX842584                244800 bp      DNA      circular BCT 21-NOV-2003
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 13/13.
ACCESSION  BX842584 AL021426 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864
VERSION    Z94121 Z97188
KEYWORDS   BX842584.1 GI:38490370
SOURCE     complete genome.
ORGANISM   Mycobacterium tuberculosis H37Rv
            Mycobacterium tuberculosis H37Rv
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrall,B.G.
TITLE      Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
JOURNAL    Nature 393 (6685), 537-544 (1998)
MEDLINE    98295987
PUBMED     9634230
REFERENCE  2
AUTHORS    Camus,J.C., Pryor,M.J.J., Medigue,C. and Cole,S.T.
TITLE      Re-annotation of the genome sequence of Mycobacterium tuberculosis
            H37Rv
JOURNAL    Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
MEDLINE    22255591
PUBMED     12368430
REFERENCE  3 (bases 1 to 244800)
AUTHORS    Parkhill,J.
TITLE      Direct Submission
SUBMITTED  (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75274 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
            On or before Nov 21, 2003 this sequence version replaced
            gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648,
            gi:3261687, gi:3261736, gi:3261805.
NOTES:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).
FEATURES   source
            1. .244800
               /organism="Mycobacterium tuberculosis H37Rv"
               /mol_type="genomic DNA"
               /strain="H37Rv"
               /db_xref="taxon:83332"
               complement(89..1396)
               /locus_tag="Rv3722c"
               complement(89..1396)
               /locus_tag="Rv3722c"
               /functions="UNKNOWN"
               /note="Rv3722c, (MTV025.070c), len: 435 aa. Conserved
               hypothetical protein, equivalent to O69513|MLCB2407.14
               (alias Q9CB73|ML2336, 463 aa) HYPOTHETICAL 46.8 KDA
               PROTEIN from Mycobacterium leprae (426 aa), FASTA scores:
               opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa
               overlap). Also highly similar to Q9RU17|DR1579 CONSERVED
               HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452
               aa), FASTA scores: opt: 1162, E(): 3.1e-67, (44.8%
               identity in 422 aa overlap); and partially similar to
               Q91371|PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas
               aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078,
               (25.85% identity in 348 aa overlap) and other
               aminotransferases. Tbpase score is 0.900. N-terminus
               extended since first submission (previously 408 aa)."

```

```

/codon_start=1
/evidence=experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAAL8044.2"
/db_xref="GI:38490371"
/translation="MSFDSLSPQELAAALHARHOODYAALQCMKALDLTRGKPSAEQL
DUSNQLLSLPGGDDYRPEGDTDRNYGGHGLPGLRAIFAEGLGIAVNLNAGNSSLE
LHMDIVAFSLYGVGDSPRFMIQEQDGIKFLCFVPGVYDRHFAITETGIMIPIMLQ
DGDVDLIEELVADVPAIKGMVTPVFGNPSGVYSWETVRLVQMTAAADPDLPLWD
NAVAVHTLTDIPEQDVNLGLAAKAGNPNRPYVFASTSKITFAGGVSFTFGGSLGNIA
WYLOYAGKKSIGCDKNQLRHLRPFQDAGVRLHMLRHQOILAKPKFALVAEVLDRIS
ESKIASWTEPKGTFISLDVLPGTARTVALAKDVGIAVTEAGASFYRKDPDDKNIR
IAPSPFSPVDNRNAVGLATCALLAATETLLNQLGLASSAPNVR"
/gene      1613..1698
/gene="trna-Ser (GGA)"
/gene="trna-Ser (GGA)"
/product="trna-Ser"
/notes="codon recognized: UCC"
/anticodon=(pos:1647..1649,aa:Ser)
1804..2568
/locus_tag="Rv3723"
1804..2568
/locus_tag="Rv3723"
/function="UNKNOWN"
/notes="Rv3723, (MTV025.071), len: 254 aa. Probable
conserved transmembrane protein, with hydrophobic
stretchches at the N-terminus, and equivalent to
O69512|ML2337|MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from
Mycobacterium leprae (250 aa), FASTA scores: opt: 1029,
E(): 1.2e-44, (64.45% identity in 253 aa overlap). Tbpase
score is 0.900."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAAL8045.1"
/db_xref="GI:2960147"
/translation="MGRKAVLWHSFSGAGVLYFYVFLPRWPELMGDTGSHSLGTGL
VMTLSDAAGOMLFGIYCAAAIAIVLGFPGFVLFVLAELPPPKPLPKPKORRLRR
KTKAGDEAPEAEAEANTELAQDEEAVEAPPESIESFGGEPESATREAPAEETA
TAEPRGUGLRNRRPTGKTSRRRRRSGVQVAKVDE"
2735..2977
/gene="cut5a"
/locus_tag="Rv3724a"
2735..2977
/gene="cut5a"
/locus_tag="Rv3724a"
/EC_number="3.1.1.-"
/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE
STRUCTURE OF PLANT CUTICLE)."
/notes="Rv3724a, (MTV025.072), len: 80 aa. Probable cut5a,
truncated cutinase precursor (EC 3.1.1.-), similar to
N-terminal end of others e.g. Q9KK87 SERINE ESTERASE
CUTINASE from Mycobacterium avium (220 aa), FASTA scores:
opt: 202, E(): 1.5e-06, (56.45% identity in 62 aa
overlap); Q9XB09|RVD2-RV1758 PROTEIN (FRAGMENT) from
Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200,
E(): 1.5e-06, (61.4% identity in 57 aa overlap); and
Q00298|CUTI_BOTCI|CUTA CUTINASE PRECURSOR from Botrytis
cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores:
opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap).
Also highly similar to others from Mycobacterium
tuberculosis e.g.
O06318|CUT3_MYCTU|Rv3451|MT3557|MTCY13E12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E():
1.2e-05, (58.0% identity in 50 aa overlap);
Q50664|CUT2_MYCTU|Rv2301|MT2358|MTCY339.08c PROBABLE
CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E():
0.00015, (59.2% identity in 49 aa overlap);
O06793|Rv1758|MTCY28.24|Z95890 HYPOTHETICAL 17.9 KDA
PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29,

```

(57.2% identity in 166 aa overlap);
O06319|RV3452|MTV13E12.05; and U00015.11 from Mycobacterium
leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase
ORF continues as RV3724B|CUT5B, frameshifting could occur
near position 4169668. Sequence has been checked but no
errors found.

/codon_start=1
/transl_table=1
/product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A"
/protein_id="CAE55632.1"
/db_xref="GI:38490372"
/translation="MDVIRWARLAVVAGTAAAVTTPLGLSAHVPMVMSABPCPDVEVV
FAGTGPPGIGSVGGFLVDALRFPQWQVTRGLRR"

gene

/gene="cut5b"

/locus_tag="RV3724B"

2874..3437

/gene="cut5b"

/locus_tag="RV3724B"

/EC_number="3.1.1.-"

/function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE
STRUCTURE OF PLANT CUTICLE)."

/notes="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b,
truncated cutinase (EC 3.1.1.-), similar to C-terminal end
of others e.g. OX8B09|RVD2-RV1758 PROTEIN (FRAGMENT) from
Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 335,
E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9K87
SERINE ESTERASE CUTINASE from Mycobacterium avium (220
aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05%
identity in 168 aa overlap). Also similar to proteins from
Mycobacterium tuberculosis e.g. O06793|RV1758|MTCV28.24
HYPOTHETICAL 17.9 KDA PROTEIN (174 aa) FASTA scores: opt:
641, E(): 2.5e-29, (57.25% identity in 166 aa overlap);
O06319|RV3452|MTV13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN
(226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%
identity in 165 aa overlap);
O06318|CUT3_MVCTU|RV3451|MT3557|MTCV13E12.04 PROBABLE
CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 307, E():
1.9e-10, (40.7% identity in 167 aa overlap);
Q10837|CUT1_MVCTU|RV1984c|MT2037|MTCV39.35 PROBABLE
CUTINASE PRECURSOR (217 aa), FASTA scores: opt: 261, E():
6.7e-08, (50.9% identity in 169 aa overlap); etc; and
U00015.11 from Mycobacterium lepra. 5'-end of gene is
RV3724B|CUT5A; frameshifting may occur near position
4169668. TParse score is 0.918."

/codon_start=1

/transl_table=1

/product="PROBABLE CUTINASE [SECOND PART] CUT5B"

/protein_id="CAE55633.1"

/db_xref="GI:38490373"

/translation="MARGSHLVLAASEDCSTHCVSQVGAKSGLGVAVNYPASNDPAS
SDFPKTVIDIRGDSHQISWAMSCPTQVGLGYSQGAAGVGVTSVAVPPAVPQOA
VPAPMAPEVHVAATVLFQAPSAQFLGQYAPPIATGPLYQPKTLQLCADGDSICGD
GNSPVAHGLYAVNGVWGQGANFAASRL"

Alignment Scores:
Pred. No.: 165 Length: 244800
Score: 108.00 Matches: 46
Percent Similarity: 35.42% Conservative: 22
Best Local Similarity: 23.96% Mismatches: 52
Query Match: 10.39% Indels: 72
DB: 1 Gaps: 8

US-10-691-383-2_COPY_435_632 (1-198) x BX842584 (1-244800)

Qy 4 GlyThrSerHisTyr-----PheArgLeuIleGlyAla 14
Db 103548 GGACTCTCCCACTTCGGCGAGCACTGCATCGTTGGCTGATCCTGGCGCTGCTGGCGCG 103489
Qy 15 AlaGluLeuAlaGlnAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArg 34
Db 103488 ATAGCCCTGCCACCGCGCGCG-----CGGAATGCTTGTG----- 103453

Qy 35 ProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 54
Db 103452 ---GCGGGGGCGGGCGGTTCGTGGCCCATGCC----- 103417
Qy 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 74
Db 103416 -----GTGCTGATCAAGCGGCTGGTGGCG-----CGT 103390
Qy 75 GlnAsnProAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 94
Db 103389 CAGCGCGCGGATCATCCGGCCATC-----GCGGTCAACGTGGACACGCCA 103345
Qy 95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
Db 103344 AGTCAACTGAGCTTTCGTGGGCACACGCCACCTCGACACCGCGCGGCTGCTCATG 103285
Qy 115 LysAlaLeuIleGlyLeuAspArgGlyGluCysPheProAsnProValPheProSer 134
Db 103284 GGCAGAGCCACCGGGCTG-----CCGCTACCGGTTGTG----- 103252
Qy 135 AspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyrGluGlyGluIle 154
Db 103252 ----- 103252
Qy 155 AsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArgPhe 174
Db 103251 -----CTGGTGGCGCGCGATGGCGCTGTCGCGAATACTGCTGGGGGTCCACTACCCCACT 103198
Qy 175 AspGlyIleGlnGlyLeuLeuLeuGlyGluThrIle 186
Db 103197 GATGTGGCGCGTGGGTGTGTTGCTCTGGCGCGCACCGGTC 103162

Search completed: September 19, 2004, 07:32:13

Job time : 2540.17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:52:28 ; Search time 8.6087 Seconds
(without alignments)
1187.397 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFQTHYFRLIGAELAQR.....GLLGETITVTLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	676	3	US-09-151-189-2
2	1039	100.0	676	4	US-09-596-794-2
3	122	11.7	22	3	US-09-151-189-9
4	122	11.7	22	4	US-09-596-794-9
5	110	10.6	21	4	US-09-596-794-11
6	108	10.4	19	3	US-09-151-189-10
7	108	10.4	19	4	US-09-596-794-10
8	99	9.5	19	3	US-09-151-189-11
9	87.5	8.4	605	4	US-09-832-496-2
10	87.5	8.4	605	4	US-09-832-616-2
11	87	8.4	604	4	US-09-832-615-2
12	87	8.4	604	4	US-09-832-617-2
13	85.5	8.2	609	3	US-09-271-778-1
14	85.5	8.2	609	4	US-09-788-871-1
15	85.5	8.2	615	4	US-09-252-991A-28135
16	83.5	8.0	600	2	US-08-679-405-2
17	83.5	8.0	600	2	US-08-842-799-2
18	83.5	8.0	600	3	US-09-271-778-2
19	83.5	8.0	600	4	US-09-788-871-2
20	83.5	8.0	600	5	FCT-US96-11458-2
21	82.5	7.9	290	4	US-09-134-001C-3273
22	81.5	7.8	599	4	US-09-832-441-2
23	81.5	7.8	599	4	US-09-833-102-2
24	80.5	7.7	386	4	US-09-732-615-4
25	78.5	7.6	480	4	US-09-252-991A-32628
26	78.5	7.6	697	4	US-09-489-039A-7485
27	75.5	7.3	244	3	US-08-750-145A-24

Sequence 28, Appl
Sequence 28, Appl
Sequence 28, Appl
Sequence 246, App
Sequence 27619, A
Sequence 13, Appl
Sequence 33097, A
Sequence 2, Appl
Sequence 12, Appl
Sequence 30391, A
Sequence 17417, A
Sequence 6983, Ap
Sequence 18842, A
Sequence 4962, Ap
Sequence 5493, Ap
Sequence 28726, A
Sequence 5912, Ap

75.5 7.3 244 3 US-08-975-698A-28
75.5 7.3 244 3 US-09-417-090-28
75.5 7.3 244 4 US-09-727-578-28
75.5 7.3 447 4 US-08-311-731A-246
75.5 7.3 502 4 US-09-252-991A-27619
75.5 7.2 308 3 US-08-952-889A-13
74.5 7.2 447 4 US-09-252-991A-33097
74.5 7.2 614 4 US-09-832-498-2
74.5 7.2 614 4 US-09-832-614A-2
74 7.1 474 4 US-09-732-615-12
74 7.1 915 4 US-09-252-991A-30391
73.5 7.1 421 4 US-09-252-991A-17417
73.5 7.1 665 4 US-09-328-352-6983
73 7.0 489 4 US-09-252-991A-18842
73 7.0 872 4 US-09-543-681A-4962
73 7.0 908 4 US-09-543-681A-5493
72.5 7.0 439 4 US-09-252-991A-28726
72.5 7.0 567 4 US-09-328-352-5912

ALIGNMENTS

RESULT 1

US-09-151-189-2
; Sequence 2, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151.189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-09-151-189-2

Query Match 100.0%; Score 1039; DB 3; Length 676;
Best Local Similarity 100.0%; Pred. No. 1.6e-119;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VNFQTHYFRLIGAELAQRSCYQKQVHRFARPEALGGTLHNTIAGDLDDADFDISLLE	60
Db	435	VNFQTHYFRLIGAELAQRSCYQKQVHRFARPEALGGTLHNTIAGDLDDADFDISLLE	494
Qy	61	NDELLKRVAEINAAQNPNVTVLLPQAIQVGSPTSPSGHATQNGAFATVLKALIGL	120
Db	495	NDELLKRVAEINAAQNPNVTVLLPQAIQVGSPTSPSGHATQNGAFATVLKALIGL	554
Qy	121	DRGECFPNPFVDDGLELINFEGACLTVEGEINKLVANVAFQRMGLIHRYDFGIQGL	180
Db	555	DRGECFPNPFVDDGLELINFEGACLTVEGEINKLVANVAFQRMGLIHRYDFGIQGL	614
Qy	181	LLGETITVTLHQLMTFF 198	
Db	615	LLGETITVTLHQLMTFF 632	

RESULT 2

US-09-596-794-2
; Sequence 2, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases and Their Uses

```
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596,794
; CURRENT FILING DATE: 2000-06-19
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-09-596-794-2

Query Match      100.0%; Score 1039; DB 4; Length 676;
Best Local Similarity 100.0%; Pred. No. 1.6e-119;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNFQTSHYFLRIGAAELAQASQYQKWQVHRFARPEALGCTLHNTIAGDLADFDISLLE 60
DB 435 VNFQTSHYFLRIGAAELAQASQYQKWQVHRFARPEALGCTLHNTIAGDLADFDISLLE 494
QY 61 NDELLKRVASINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATQNGAPATVLKALIGL 120
DB 495 NDELLKRVASINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATQNGAPATVLKALIGL 554
QY 121 DRGCECFPNVFPDDGLELINFEGACLTVEGEINKLVANVAFGRQMLGIHYRFDGIQGL 180
DB 555 DRGCECFPNVFPDDGLELINFEGACLTVEGEINKLVANVAFGRQMLGIHYRFDGIQGL 614
QY 181 LLGETITVRLHQLMTF 198
DB 615 LLGETITVRLHQLMTF 632

RESULT 3
US-09-151-189-9
; Sequence 9, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151,189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473
US-09-151-189-9

Query Match      11.7%; Score 122; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.7e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AQRASYQKWQVHRFARPEALG 39
DB 1 AQRASYQKWQVHRFARPEALG 22

RESULT 4
US-09-596-794-9
; Sequence 9, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: 023070-087100US
; CURRENT FILING DATE: 2000-06-19
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609,
US-09-596-794-11

Query Match      10.6%; Score 110; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.8e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 NKLAVNVAFGRQMLGIHYRFD 175
DB 1 NKLAVNVAFGRQMLGIHYRFD 21

RESULT 6
US-09-151-189-10
; Sequence 10, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
```

```
; CURRENT APPLICATION NUMBER: US/09/151.189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546
US-09-151-189-10

Query Match      10.4%; Score 108; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  94 PTHPSYPSGHATONGAFAT 112
Db  1 PTHPSYPSGHATONGAFAT 19

RESULT 7
US-09-596-794-10
; Sequence 10, Application US/09596794
; Patent No. 6656715
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/09/596.794
; CURRENT FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546,
; OTHER INFORMATION: 2nd conserved motif
US-09-596-794-10

Query Match      10.4%; Score 108; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  94 PTHPSYPSGHATONGAFAT 112
Db  1 PTHPSYPSGHATONGAFAT 19

RESULT 8
US-09-151-189-11
; Sequence 11, Application US/09151189
; Patent No. 6232457
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/151.189
; CURRENT FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 19
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609
US-09-151-189-11

Query Match      9.5%; Score 99; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  157 LAVNVAFGRQMLGIHYRFD 175
Db  1 LAVNVAFGRQMLGIHYRFD 19

RESULT 9
US-09-832-496-2
; Sequence 2, Application US/09832496
; Patent No. 6503508
; GENERAL INFORMATION:
; APPLICANT: Danielssen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10042.200-US
; CURRENT APPLICATION NUMBER: US/09/832.496
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Dreschlera hartliebii
US-09-832-496-2

Query Match      8.4%; Score 87.5; DB 4; Length 605;
Best Local Similarity 23.2%; Pred. No. 0.1;
Matches 32; Conservative 17; Mismatches 38; Indels 51; Gaps 5;

Qy  71 INAAQNPNNVETLLPQAIQVGSPTSPSPSGHATONGAF----- 110
Db  376 LGAPATNTDIPF-----KPPFPAYSGHATFGAVFQVRRYNGRVGTWDDP 426

Qy  111 -----ATVLKALIGLDRGECFPNPVPFSDDGLELI-----NFGAC-LTYEGEINKL 157
Db  427 DNTAIDMVVSEELNGLSRDLQRDYDPTAPIEDQPIVTRVVRHFNLSAWELMFENAIISRI 486

Qy  158 AVNVAFGRQMLGIHYRFD 175
Db  487 -----FLGVHWRFD 495

RESULT 10
US-09-832-616-2
; Sequence 2, Application US/09832616
; Patent No. 6506586
; GENERAL INFORMATION:
; APPLICANT: Danielssen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10175.200-US
; CURRENT APPLICATION NUMBER: US/09/832.616
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Dreschlera hartliebii
US-09-832-616-2

Query Match      8.4%; Score 87.5; DB 4; Length 605;
Best Local Similarity 23.2%; Pred. No. 0.1;
```


Matches 32; Conservative 17; Mismatches 38; Indels 51; Gaps 5;
Qy 71 INAAQPNNEVTYLLPQAIQVGSPTHPSPYSGHATQNGAF----- 110
Db 376 LGAPATNTNDIPF-----KPPFPAYPSGHATFGSAVQVQVRRYNGRVGTWKDDEP 426
Qy 111 -----ATVLKALIGLDRGCECFNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 157
Db 427 DNTAIDMWVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI 486
Qy 158 AVNVAFGROMLGIHYRFD 175
Db 487 -----FLGVHWRFD 495

RESULT 11

US-09-832-615-2
; Sequence 2, Application US/09832615
; Patent No. 6509181
; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Polypeptides having haloperoxidase activity
; FILE REFERENCE: 10041.200-US
; CURRENT APPLICATION NUMBER: US/09/832,615
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Dendryphiella salina
US-09-832-615-2

Query Match 8.4%; Score 87; DB 4; Length 604;
Best Local Similarity 22.4%; Pred. No. 0.12;
Matches 43; Conservative 21; Mismatches 52; Indels 76; Gaps 8;

Qy 51 DADFDISLLENDELLKRVAEINAAQ-----NPNNEVTYLLPQA----- 88
Db 318 EADLATSEVNADFARLLALNVASADAGIFSWKEWEPYRPLSGVRDGRPDHADPF 377
Qy 89 -IQVGSPT-----HPSYPSGHAT-----QNGAFAT----- 112
Db 378 WLTLGAPATNTNDIPFKPPFPAYPSGHATFGSAVQVQVRRYNGRVGTWKDDEPNDIAID 437
Qy 113 --VLKALIGLDRGCECFNPVPFSDGGLLELI-----NFGAC-LTYEGEINKLNVAV 163
Db 438 MVVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI----- 491
Qy 164 GROMLGIHYRFD 175
Db 492 ---FLGVHWRFD 500

RESULT 12

US-09-832-617-2
; Sequence 2, Application US/09832617
; Patent No. 6511835
; GENERAL INFORMATION:
; APPLICANT: Danielsen, Steffen
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic acids encoding polypeptides having haloperoxidase activity
; FILE REFERENCE: 10174.200-US
; CURRENT APPLICATION NUMBER: US/09/832,617
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 604
; TYPE: PRT
; ORGANISM: Dendryphiella salina
US-09-832-617-2

Query Match 8.4%; Score 87; DB 4; Length 604;
Best Local Similarity 22.4%; Pred. No. 0.12;
Matches 43; Conservative 21; Mismatches 52; Indels 76; Gaps 8;
Qy 51 DADFDISLLENDELLKRVAEINAAQ-----NPNNEVTYLLPQA----- 88
Db 318 EADLATSEVNADFARLLALNVASADAGIFSWKEWEPYRPLSGVRDGRPDHADPF 377
Qy 89 -IQVGSPT-----HPSYPSGHAT-----QNGAFAT----- 112
Db 378 WLTLGAPATNTNDIPFKPPFPAYPSGHATFGSAVQVQVRRYNGRVGTWKDDEPNDIAID 437
Qy 113 --VLKALIGLDRGCECFNPVPFSDGGLLELI-----NFGAC-LTYEGEINKLNVAV 163
Db 438 MVVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI----- 491
Qy 164 GROMLGIHYRFD 175
Db 492 ---FLGVHWRFD 500

RESULT 13

US-09-271-778-1
; Sequence 1, Application US/09271778
; Patent No. 6221821
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Jorgensen, Louise
; TITLE OF INVENTION: Haloperoxidases With Altered pH Profiles
; FILE REFERENCE: 5516.200-US
; CURRENT APPLICATION NUMBER: US/09/271,778
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: PA 1998 00374
; EARLIER FILING DATE: 1998-03-18
; EARLIER APPLICATION NUMBER: 60/079,228
; EARLIER FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Curvularia inaequalis
US-09-271-778-1

Query Match 8.2%; Score 85.5; DB 3; Length 609;
Best Local Similarity 21.7%; Pred. No. 0.18; 35; Indels 51; Gaps 5;
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;

Qy 71 INAAQPNNEVTYLLPQAIQVGSPTHPSPYSGHATQNGAFATVLK----- 115
Db 381 LGAPATNTNDIPF-----KPPFPAYPSGHATFGSAVQVQVRRYNGRVGTWKDDEP 431
Qy 116 -----ALIGLDRGCECFNPVPFSDGGLLELI-----NFGAC-LTYEGEINKL 157
Db 432 DNTAIDMWVSEELNGLSRDLRQYDPTAPIDQPGIVRTRVVRHFSAWELMFENAIISRI 491
Qy 158 AVNVAFGROMLGIHYRFD 175
Db 492 -----FLGVHWRFD 500

RESULT 14

US-09-788-871-1
; Sequence 1, Application US/09788871
; Patent No. 6372465
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Jorgensen, Louise
; TITLE OF INVENTION: Haloperoxidases With Altered pH Profiles
; FILE REFERENCE: 5516.200-US
; CURRENT APPLICATION NUMBER: US/09/788,871
; CURRENT FILING DATE: 2001-02-20

```
; PRIOR APPLICATION NUMBER: 09/271,778
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 609
; TYPE: PRT
; ORGANISM: Curvularia inaequalis
US-09-788-871-1

Query Match      8.2%  Score 85.5; DB 4; Length 609;
Best Local Similarity 21.7%; Pred.No.0.18; 35; Indels 51; Gaps 5;
Matches 30; Conservative 22; Mismatches 22;

Qy 71 INAAQPNNEVTYLLPOAIQVSGPTSPSHGATONGAFATVLK-----115
Db 381 LGAPATNTDIPF-----KPPFPAYPSGHATFGAVFQVRRYYNGRVGTWKDEP 431
Qy 116 -----ALGLDRGCEFPNPVPSDDGLELI-----NFGAC-LTYEGEINKL 157
Db 432 DNIAIDMMISEELNGVNRDLRQPYDPTAPIEDQGVIVRIVRHFDSEAWELMFENAI SRI 491
Qy 158 AVNVAFGRQMLGHIHYRED 175
Db 492 -----FLGVHWRED 500

RESULT 15
US-09-252-991A-28135
; Sequence 28135, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28135
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28135

Query Match      8.2%  Score 85.5; DB 4; Length 615;
Best Local Similarity 22.3%; Pred.No.0.18;
Matches 42; Conservative 20; Mismatches 65; Indels 61; Gaps 7;

Qy 17 LAQRASCYQKQVHRFARPEA-----LGGTLHTNTIAGDLDDADFDISLLENDELL 65
Db 385 VAQRQADQDQDHR-HRLPAPEGEGQGVGEQSHLAGTLHRYVAGG-----427
Qy 66 KRVAEINAAQNPNEVTYLLPOAIQVG-----SPTHSPSPSGHATQNGAPA 111
Db 428 ---EQRAAAGEDHQVGMQRQARAEAGPGQVEVHLGPDQLRGDQYAOQPHADHAPYHGHDG 484
Qy 112 TVLKALIGLDRGCECF-----PNVPFSDDGLELINFEAGCLTYEGE-INKLAV 159
Db 485 ELANHLVVVGFRRCFVHLYSPERNEQPATAVEGGDGP-----AGACNLRAGAGVKQLQV 540
Qy 160 NVAFGRQM 167
Db 541 DMEFGYQM 548
```

Search completed: September 19, 2004, 04:00:33
Job time : 9.6087 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:55:28 ; Search time 39.4188 Seconds
(without alignments)
1613.049 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFGTSHYFRLIGAELAQR.....GLLGGTITVTRLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 segs, 32113274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	676	9	US-09-840-762A-2
2	1039	100.0	676	16	US-10-691-383-2
3	122	11.7	22	9	US-09-840-762A-9
4	122	11.7	22	16	US-10-691-383-9
5	110	10.6	21	16	US-10-691-383-11
6	108	10.4	19	9	US-09-840-762A-10
7	108	10.4	19	16	US-10-691-383-10
8	108	10.4	165	14	US-10-080-170-629
9	108	10.4	165	16	US-10-080-170-629
10	99	9.5	19	9	US-09-840-762A-11
11	93	9.0	525	14	US-10-156-761-15017
12	89	8.6	303	16	US-10-767-701-44002
13	89	8.6	333	12	US-10-425-114-60720
14	88.5	8.5	647	9	US-09-969-680A-16
15	88.5	8.5	647	10	US-09-969-680A-16

16	88.5	8.5	647	12	US-10-169-395-4	Sequence 4, Appli
17	88.5	8.5	647	15	US-10-353-690-76	Sequence 76, Appl
18	88.5	8.5	647	15	US-10-295-027-1337	Sequence 1337, Ap
19	87.5	8.4	605	9	US-09-832-616-2	Sequence 2, Appli
20	87.5	8.4	605	9	US-09-832-496-2	Sequence 2, Appli
21	87	8.4	198	12	US-10-282-122A-77965	Sequence 77965, A
22	86	8.3	192	12	US-10-282-122A-63648	Sequence 63648, A
23	86	8.3	192	14	US-10-080-170-23	Sequence 23, Appl
24	86	8.3	192	16	US-10-080-170-23	Sequence 23, Appl
25	86	8.3	330	12	US-10-425-114-66217	Sequence 66217, A
26	85.5	8.2	1256	9	US-09-854-356-14	Sequence 14, Appl
27	85.5	8.2	1273	12	US-10-363-616-325	Sequence 325, App
28	84	8.1	168	9	US-09-738-626-6676	Sequence 6676, Ap
29	84	8.1	453	14	US-10-156-761-8288	Sequence 8288, Ap
30	83.5	8.0	175	12	US-10-282-122A-52831	Sequence 52831, A
31	83.5	8.0	626	14	US-10-097-340-95	Sequence 95, Appl
32	83.5	8.0	626	15	US-10-093-463-120	Sequence 120, App
33	83.5	8.0	647	11	US-09-833-245-686	Sequence 686, App
34	82.5	7.9	341	12	US-10-424-599-185920	Sequence 185920,
35	81.5	7.8	599	9	US-09-833-102-2	Sequence 2, Appli
36	81.5	7.8	599	9	US-09-832-441-2	Sequence 2, Appli
37	81.5	7.8	737	14	US-10-017-161-2394	Sequence 2394, Ap
38	81.5	7.8	737	15	US-10-292-798-2036	Sequence 2036, Ap
39	81	7.8	509	16	US-10-437-963-131715	Sequence 131715,
40	80.5	7.7	330	12	US-10-440-503-210	Sequence 210, App
41	80.5	7.7	330	12	US-10-461-925-210	Sequence 210, App
42	80.5	7.7	330	14	US-10-146-772-210	Sequence 210, App
43	80.5	7.7	330	15	US-10-241-742-210	Sequence 210, App
44	80.5	7.7	330	15	US-10-440-523-210	Sequence 210, App
45	80.5	7.7	386	14	US-10-273-051-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-840-762A-2
; Sequence 2, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840, 762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
; US-09-840-762A-2

Query Match	100.0%;	Score 1039;	DB 9;	Length 676;
Best Local Similarity	100.0%;	Pred. No. 1.5e-106;		
Matches 198;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	VNFGTSHYFRLIGAELAQRASCYQKWQVHRFAPEALGGLTHTNTIAGDLDDADFDISLLE	60	
Db	435	VNFGTSHYFRLIGAELAQRASCYQKWQVHRFAPEALGGLTHTNTIAGDLDDADFDISLLE	494	
Qy	61	NDELLKRVAEINAAONPNNEVTYLLPOAIOVGSPHTPSYSGHATONGAFATVUKALIGL	120	
Db	495	NDELLKRVAEINAAONPNNEVTYLLPOAIOVGSPHTPSYSGHATONGAFATVUKALIGL	554	
Qy	121	DRGCECFNPFVDDGGLLEINFEGACLTVEGEINKLVANVAFQROMLGHYRFDGQGL	180	
Db	555	DRGCECFNPFVDDGGLLEINFEGACLTVEGEINKLVANVAFQROMLGHYRFDGQGL	614	

```
QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

RESULT 2
US-10-691-383-2
; Sequence 2, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; PRIOR FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Fucus distichus
US-10-691-383-2

Query Match 100.0%; Score 1039; DB 16; Length 676;
Best Local Similarity 100.0%; Pred. No. 1.5e-106;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VNFGTSHYFLIGAAELAQASCYQKQVHRFARPEALGTLHTNTIAGDLADFDISLLE 60
Db 435 VNFGTSHYFLIGAAELAQASCYQKQVHRFARPEALGTLHTNTIAGDLADFDISLLE 494

QY 61 NDELLKRVAEINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATONGAFATVKKALIGL 120
Db 495 NDELLKRVAEINAAQNPNNEVTYLLPOAIOVGSPTSPSGHATONGAFATVKKALIGL 554

QY 121 DRGCECFNPFVPSDDGLELINFEGACLTYEIGINKLAVNVAFGQRLGIHYRFDGIQGL 180
Db 555 DRGCECFNPFVPSDDGLELINFEGACLTYEIGINKLAVNVAFGQRLGIHYRFDGIQGL 614

QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

RESULT 3
US-09-840-762A-9
; Sequence 9, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473

QY 181 LLGETITVRTLHQLMTF 198
Db 615 LLGETITVRTLHQLMTF 632

Query Match 11.7%; Score 122; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AQRASYCQKQVHRFARPEALG 39
Db 1 AQRASYCQKQVHRFARPEALG 22

RESULT 4
US-10-691-383-9
; Sequence 9, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved Fucus
; OTHER INFORMATION: vanadium-binding region 1, amino acids 452-473,
; OTHER INFORMATION: 1st conserved motif

US-10-691-383-9

Query Match 11.7%; Score 122; DB 16; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AQRASYCQKQVHRFARPEALG 39
Db 1 AQRASYCQKQVHRFARPEALG 22

RESULT 5
US-10-691-383-11
; Sequence 11, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:conserved Fucus
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609,
```

```
; OTHER INFORMATION: 3rd conserved motif
US-10-691-383-11

Query Match      10.6%; Score 110; DB 16; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.5e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 NKLAVNVAFGRQMLGIHYRF 175
Db 1 NKLAVNVAFGRQMLGIHYRF 21

RESULT 6
US-09-840-762A-10
; Sequence 10, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: Ng, Kwan L.
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 2001-04-23
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546
US-09-840-762A-10

Query Match      10.4%; Score 108; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATQNGAFAT 112
Db 1 PTHPSYPSGHATQNGAFAT 19

RESULT 7
US-10-691-383-10
; Sequence 10, Application US/10691383
; Publication No. US20040110260A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Minimal Catalytic Vanadium Haloperoxidases
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 023070-087110US
; CURRENT APPLICATION NUMBER: US/10/691,383
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: US/09/596,794
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: US 09/151,189
; PRIOR FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved Fucus
; OTHER INFORMATION: vanadium-binding region 2, amino acids 528-546,
; OTHER INFORMATION: 2nd conserved motif
US-10-691-383-10

Query Match      10.4%; Score 108; DB 16; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 PTHPSYPSGHATQNGAFAT 112
Db 1 PTHPSYPSGHATQNGAFAT 19

RESULT 8
US-10-080-170-629
; Sequence 629, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 629
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-629

Query Match      10.4%; Score 108; DB 14; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.0014;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

Qy 4 GTSHY-----FRLIGAAELAQASCYQKQVHRFAEALGGTLHNTIAGDLDAF 54
Db 20 GLSHFGEHCITGWLITALLGAIALPRRR---REWLIV-----AGAGAFVAHAIA----- 63
Qy 55 DISLENDELLKVAEINNAQNPNNEVYLLPQAIQVGSPTSPSYPSGHATQNGAFATVL 114
Db 64 -----VLTKRLVR---RQPDHPAI-----AVNVDTPSQSFPSAHATSTTAAALLM 107
Qy 115 KALITGLDRGECFPNPVPFSDGGLLEINPEGCACITYEINKLAVNVAFGRQMLGIHYRF 174
Db 108 GRATGL-----PLPVV-----LVPPNALSRIILGVHYP 136
Qy 175 DGIQGLLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 9
US-10-080-170-629
; Sequence 629, Application US/10080170
; Publication No. US20040121322A9
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 629
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
```

US-10-080-170-629

Query Match	10.4%	Score 108;	DB 16;	Length 165;		
Best Local Similarity	24.0%;	Pred. No. 0.0014;				
Matches	46;	Conservative 22;	Mismatches 52;	Indels 72; Gaps 8;		
Qy	4	GTSHY-----	FRLIGAAELAQRASCYKQWVHRFARPEALGGTGLHNTIAGDLQADF	54		
		:	:	:		
Db	20	GLSHFGHCIGTWLIALGAIALPRRR-----	RENLV-----	AGAGAFVAHAIA-----	63	
		:	:	:		
Qy	55	DISLENDELLKRVAEINAAQNPNVETLLPQAIQVGSPTHSPVPSGATHONGAFATVL	114			
		:	:	:	:	
Db	64	-----	VLIKRLVR--	RQRDPHPAI-----	AVNVDTPSQLSPFSAHATSTTAAAILM	107
		:	:	:	:	
Qy	115	KALIGDRGCECFNPVPFPPSDGLELINFEGACLTYEGINKLVAVNVAFGOMLGIHYRF	174			
		:	:	:	:	
Db	108	GRATGL-----	PLPVV-----	-----	LVPMPALSILLOVHYPS	136
		:	:	:	:	
Qy	175	DGIQGLLGETI	186			
		:	:	:	:	
Db	137	DVAVGVALGATV	148			

RESULT 10

```

US-09-840-762A-11
; Sequence 11, Application US/09840762A
; Patent No. US20020035245A1
; GENERAL INFORMATION:
; APPLICANT: Vreeland, Valerie
; APPLICANT: NG, Kwan L.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Recombinant Vanadium Haloperoxidases and Their Uses
; FILE REFERENCE: 023070-087100US
; CURRENT APPLICATION NUMBER: US/09/840,762A
; CURRENT FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 09/151,189
; PRIOR FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 19
; TYPE: PRP
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: conserved
; OTHER INFORMATION: vanadium-binding region 3, amino acids 591-609
US-09-840-762A-11

```

```

Query Match      9.5%; Score 99; DB 9; Length 19;
Best Local Similarity 100.0%; Pred. No. 0.00066;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      157 LAVNVAFGRQMLGIHYRFD 175
      |||||
Db      1 LAVNVAFGRQMLGIHYRFD 19

```

RESULT 11

```

US-10-156-761-15017
; Sequence 15017, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29

```

```

; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15017
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15017

```

```

Query Match      9.0%; Score 93; DB 14; Length 525;
Best Local Similarity 27.7%; Pred. No. 0.34;
Matches 28; Conservative 12; Mismatches 31; Indels 30; Gaps 3;

Qy      84  LLPQAIVGSPTH-PSYPGSHATONGAFATVTKALIGLDRGGCFPNVPFSDGLELIN 142
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      90  LVPQIRHLTRQPHITTFSPSGHSAFAATG-----VALESTG----- 127

Qy      143  FEGACLTYEIEINKLVANVAFGRQMLGIHYRFDGICQLLIG 183

Db      128  -----YGLVAPLAAAVAFSRVYGVHYGPGDVLAMG 161

```

RESULT 12

```

US-10-767-701-44002
; Sequence 44002, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: KOVALIC, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 44002
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C9661_1.pep
US-10-767-701-44002

```

Query Match	8.8%;	Score 89;	DB 16;	Length 303;
Best Local Similarity	24.7%;	Pred. No. 0.44;		
Matches 54;	Conservative 24;	Mismatches 61;	Indels 80;	Gaps 13;
Qy	11	LIGAAELAQR-ASCYQKQVHREA-RPEALGGTGLHNTIAGDLADADFDISLLENDELK-R	67	
Db	21	LACASAAARGASGFPSRVVQLSWRPRAF--LHKGFLSDAEC-DHLIVLAKDKLEKSM	76	
Qy	68	VAEINAAQNPNNEV-----TYLLP---QAIV-----GS	93	
Db	77	VADNESGKSVQSEVRTSSGFMLEKKODEVVRGIEERIAAWTFLPPENGESIQILHYONGE	136	
Qy	94	PTPSPSYSGHATONGA-----FATVLKALIGLDRGGE-CPNP-----VFFSDD-----	136	
Db	137	KYBEHYDYFDKKNQALCGHRIATVLMYLSNVKGGSETIFPNAEGKLLQPKDITWSDCAR	196	
Qy	137	-----GLELINF-----EGACLTVEGE	153	
Db	197	NGVAVKPKVGDAALLFFSLHPDATTDSBSLHSGSCPVIEGQ	235	

RESULT 13

US-10-425-114-60720
; Sequence 60720, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 60720
LENGTH: 333
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3587-278-Cl1_FLI.pbp
US-10-425-114-60720

Query Match 8.6%; Score 89; DB 12; Length 333;
Best Local Similarity 24.3%; Pred. No. 0.5;
Matches 53; Conservative 23; Mismatches 64; Indels 78; Gaps 11;

Qy 11 LIGAAELAQR-ASCYQKQVHRFA-RPEALGGTTLHNTIAG--DLDADFDISLENDE--L 64
Db 51 LTGCASAASRGAGSFDPSRVVQLSWRPRAF---LHKGFLLDAECDHLIALAKOKLEKSMV 107
Qy 69 AETNAAGQPNNEV-----TVLLP---QAIVQ-----GSP 94
Db 108 ADNKGSKSVQSEVTRTSSGMELEKKQDEVTVTRIERISAWTFPENGCEAIQILHYQNGEK 167
Qy 95 THPSYPSGHATQNGA-----FATVVKALIGLDGRGE-CFPNP-----VFPSSD----- 136
Db 168 YEPHYDYFHDKNQALGGHRIATVLMYLSNVKSGETIFPNAEGLQPKDDTWSDCARN 227
Qy 137 -----GLELINF-----EGACLYEGE 153
Db 228 GYAVKPKVGDALEFFSLHPDSTTDSLSHGSCPVIQEQ 265

RESULT 14
US-09-965-529-16
Sequence 16, Application US/09965529
Publication No. US20020182671A1
GENERAL INFORMATION:
APPLICANT: LAL, Preeti
APPLICANT: YUE, Henry
APPLICANT: BANDMAN, Olga
APPLICANT: TANG, Y. Tom
APPLICANT: BURFORD, Neil
APPLICANT: AZIMZAI, Valda
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LU, Dyung Aina M.
APPLICANT: PATTERSON, Chandra
TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS
FILE REFERENCE: PP-0731 USA
CURRENT APPLICATION NUMBER: US/09/965,529
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/149,641; 60/164,203; PCT/US00/22315
PRIOR FILING DATE: 1999-08-17; 1999-11-09; 2000-08-14
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PERL Program
SEQ ID NO 16
LENGTH: 647
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020182671A1 1336728CD1
US-09-965-529-16

Query Match 8.5%; Score 88.5; DB 9; Length 647;
Best Local Similarity 28.4%; Pred. No. 1.5;

Matches 38; Conservative 18; Mismatches 39; Indels 39; Gaps 7;
Qy 15 AELAQASCYQKQVHRFA-RPEALGGTTLHNTIAG--DLDADFDISLENDE--L 64
Db 94 ARLSAAAVLY-----LSNPE---GTCEPTRAGLWASHADHLLALLESPPKALTPGLSWL 143
Qy 65 LKRVAEINAAQNPN-----EVTYLLPQAIQVSGTPHPSYPSGHATQNGAFATVVKALIGLD 121
Db 144 LORMQARAAGCTKTKACVDIPQLLEAVGAGP-----GSAGGVLAALLDHHV 190
Qy 122 RGGECEP---PNPVF 132
Db 191 RSGSCFHALPSPQY 204

RESULT 15
US-09-969-680A-16
Sequence 16, Application US/09969680A
Publication No. US20030124649A1
GENERAL INFORMATION:
APPLICANT: LAL, Preeti; YUE, Henry
APPLICANT: TANG, Y. Tom; BANDMAN, Olga
APPLICANT: BURFORD, Neil; AZIMZAI, Valda
APPLICANT: BAUGHN, Mariah R.; LU, Dyung Aina M.
APPLICANT: PATTERSON, Chandra
TITLE OF INVENTION: MEMBRANE ASSOCIATED PROTEINS
FILE REFERENCE: PP-0731-1 USA
CURRENT APPLICATION NUMBER: US/09/969,680A
CURRENT FILING DATE: 2001-10-02
PRIOR APPLICATION NUMBER: US00/22315
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/149,641
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/164,203
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PERL Program
SEQ ID NO 16
LENGTH: 647
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030124649A1 1336728CD1
US-09-969-680A-16

Query Match 8.5%; Score 88.5; DB 10; Length 647;
Best Local Similarity 28.4%; Pred. No. 1.5; Mismatches 39; Indels 39; Gaps 7;
Matches 38; Conservative 18; Mismatches 39; Indels 39; Gaps 7;
Qy 15 AELAQASCYQKQVHRFA-RPEALGGTTLHNTIAG--DLDADFDISLENDE--L 64
Db 94 ARLSAAAVLY-----LSNPE---GTCEPTRAGLWASHADHLLALLESPPKALTPGLSWL 143
Qy 65 LKRVAEINAAQNPN-----EVTYLLPQAIQVSGTPHPSYPSGHATQNGAFATVVKALIGLD 121
Db 144 LORMQARAAGCTKTKACVDIPQLLEAVGAGP-----GSAGGVLAALLDHHV 190
Qy 122 RGGECEP---PNPVF 132
Db 191 RSGSCFHALPSPQY 204

Search completed: September 19, 2004, 04:03:37
Job time : 41.4188 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:11:56 ; Search time 38.0595 Seconds
(without alignments)
1469.920 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFQTHYFLRIGAEALQRA.....GLLGLTITVRLHQLMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1039	100.0	676	3	AAY87469 Fucus van
2	1039	100.0	676	5	AAE17501 Fucus dis
3	332.5	32.0	598	4	AAE05750 Corallina
4	331.5	31.9	598	4	AAE05756 Corallina
5	329.5	31.7	598	2	AAW80550 Haloperox
6	322.5	31.0	598	4	AAE05755 Corallina
7	322	31.0	597	2	AAW80551 Haloperox
8	322	31.0	598	4	AAE05757 Corallina
9	122	11.7	22	5	AAE17502 Fucus dis
10	110	10.6	21	5	AAE17504 Fucus dis
11	108	10.4	19	5	AAE17503 Fucus dis
12	108	10.4	165	5	ABU05978 M. tuberc
13	100	9.6	37	4	AAE05754 Vanadium
14	88.5	8.5	396	4	ABB70948 Drosophil
15	88.5	8.5	647	4	AAE06574 Human pro
16	88.5	8.5	647	4	AAE06574 Human mem
17	88.5	8.5	647	5	AAU74618 Oestrogen
18	88.5	8.5	647	7	ADE31719 Human 646
19	87.5	8.4	605	5	AAG77947 Haloperox
20	87.5	8.4	605	5	AAM51613 Dreschler
21	87	8.4	198	6	ABU50041 Protein e
22	87	8.4	604	5	AAU09943 Dendryphi
23	87	8.4	604	5	AAM52241 Dendryphi
24	86.5	8.3	441	2	AAY40501 A. thalia
25	86	8.3	192	5	ABU05372 M. tuberc

26	86	8.3	192	6	ABU35724 Protein e
27	85.5	8.2	609	2	AAY27441 C. inaequ
28	85.5	8.2	640	2	AAR82249 Chloroper
29	85.5	8.2	920	5	AAM51152 Mouse Her
30	85.5	8.2	926	5	AAM51153 Mouse Her
31	85.5	8.2	1256	3	AAB21206 Mouse Her
32	85.5	8.2	1256	4	AAG62860 Amino aci
33	85.5	8.2	1256	5	AAM51151 Mouse Her
34	85.5	8.2	1273	5	ABP62888 Human pol
35	85.5	8.2	1437	6	ABP98857 Human str
36	85.5	8.2	1819	7	ADE60309 Human Pro
37	84	8.1	168	4	AAG29222 C glutami
38	83.5	8.0	175	6	ABU24907 Protein e
39	83.5	8.0	600	2	AAW12042 Curvulari
40	83.5	8.0	600	2	AAY27442 C. verruc
41	83.5	8.0	622	5	ABP64821 Human pro
42	83.5	8.0	626	5	AAU74619 Oestrogen
43	83.5	8.0	626	5	ABG96322 Human ova
44	83.5	8.0	626	5	ABP70108 Human NOV
45	83.5	8.0	647	4	AAE01677 Human gen

ALIGNMENTS

RESULT 1

AAAY87469

ID AAY87469 standard; protein; 676 AA.

AC AAY87469;

XX

DT 03-JUL-2000 (first entry)

XX

DE Fucus vanadium bromoperoxidase.

XX

XX Vanadium bromoperoxidase; vanadium haloperoxidase; alga; ODA;
KW o-dianisidine oxidation; vanadium ion; halogenation; antimicrobial agent;
KW phenolic adhesive production.

XX Fucus sp.

XX

PH Key Location/Qualifiers

FT Region 441..676

FT /note= "Conserved C-terminal region"

FT Region 452..473

FT /note= "Conserved vanadium-binding region #1"

FT Region 528..546

FT /note= "Conserved vanadium-binding region #2"

FT Region 591..609

FT /note= "Conserved vanadium-binding region #3"

WO200015771-A1.

PD 23-MAR-2000.

XX 27-AUG-1999; 99WO-US019893.

XX 10-SEP-1998; 98US-00151189.

XX (REGC) UNIV CALIFORNIA.

PA Vreeland V, Ng KL;

XX WPI; 2000-271414/23.

XX N-PSDB; AAA10305.

DR

XX Fucus vanadium bromoperoxidase polypeptide and coding sequence, useful

PT for halogenating or oxidizing compounds, e.g. as antibacterial agents.

XX Claim 5; Page 42-44; 46pp; English.

PS This sequence represents a vanadium bromoperoxidase from species of

XX Fucus. This enzyme is 73.4 kD in mass and catalyses the oxidation of o-

CC

CC dianisidine (ODA) when complexed with a vanadium ion. The cDNA encoding
 CC the enzyme was isolated from a Fucus 2-cell embryo cDNA library using
 CC hybridisation probes (AA10306-A10307) corresponding to the second and
 CC third regions that are conserved between Curvularia and Ascoplyum
 CC vanadium haloperoxidase active sites. Fucus vanadium bromoperoxidase
 CC cDNAs (one full-length, two 5' - truncated) were then cloned into the pET
 CC LIC (ligation independent cloning) vector for expression in Escherichia
 CC coli. The vanadium bromoperoxidase proteins were expressed fused to an N-
 CC terminal thioedoxin tag which optimises correct folding. The Fucus
 CC vanadium bromoperoxidase has a specific activity several-fold higher than
 CC other algal vanadium haloperoxidases for which at least partial sequences
 CC have been reported. Vanadium haloperoxidases can be used in a method for
 CC enzymatically halogenating a compound, and in a method for enzymatically
 CC oxidising a compound. They can be used to halogenate various substances,
 CC including proteins. They can be used to produce epoxides from alkenes,
 CC halogenated ketones from alkenes, to produce alpha, gamma-halohydrins
 CC from cyclopropanes, and to produce dihalogenated products from alkenes
 CC and alkenes. Vanadium haloperoxidases may also be used to oxidise various
 CC compounds, making them useful in signal generating systems in place of
 CC horseradish peroxidase. The enzymes can also be used as enzymatic
 CC antimicrobial agents and in the production of phenolic adhesives
 XX
 SQ Sequence 676 AA;

Query Match 100.0%; Score 1039; DB 3; Length 676;
 Best Local Similarity 100.0%; Pred. No. 4.6e-114;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VNFSTSHYFRLIGAAELAQASCYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 60
 DB 435 VNFSTSHYFRLIGAAELAQASCYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 494
 QY 61 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQNGAFATVVKALIGL 120
 DB 495 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQNGAFATVVKALIGL 554
 QY 121 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQWLGHIHYFDPGQGL 180
 DB 555 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQWLGHIHYFDPGQGL 614
 QY 181 LLGETITVRTLHOELMTF 198
 DB 615 LLGETITVRTLHOELMTF 632

RESULT 2
 ID AAE17501
 XX AAE17501 standard; protein; 676 AA.
 XX AAE17501;

XX
 DT 07-AUG-2003 (revised)
 DT 01-MAY-2002 (first entry)
 XX
 DE Fucus distiches vanadium haloperoxidase.
 XX
 KW Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
 KW epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
 KW industrial catalysis; enzyme; catalyst.
 XX
 OS Fucus distichus.
 XX
 XX Location/Qualifiers
 FH 439..461
 FT Region /note= "Catalytic frame helices (CFH) alpha-helix 1"
 FT 469..481
 FT Region /note= "Catalytic frame helices (CFH) alpha-helix 1a
 FT (alpha1-alpha2 loop)"
 FT 490..493
 FT Region /note= "Catalytic frame helices (CFH) alpha-helix 1b
 FT (alpha1-alpha2 loop)"
 FT 496..509
 FT Region /note= "Catalytic frame helices (CFH) alpha-helix 1c

FT Region (alpha1-alpha2 loop)"
 FT 536..552
 FT /note= "Catalytic frame helices (CFH) alpha-helix 2"
 FT 554..557
 FT /note= "Catalytic frame helices (CFH) alpha-helix 2d
 FT (alpha2-alpha3 loop)"
 FT 584..602
 FT /note= "Catalytic frame helices (CFH) alpha-helix 3"
 FT 607..631
 FT /note= "Catalytic frame helices (CFH) alpha-helix 4"
 XX WO200200838-A2.
 PN 03-JAN-2002.
 XX
 XX 07-JUN-2001; 2001WO-US018602.
 XX
 XX 19-JUN-2000; 2000US-00596794.
 XX (REGC) UNIV CALIFORNIA.
 XX Vreeland V;
 XX
 XX WPI; 2002-154651/20.
 DR N-PSDB; AAD28155.
 DR
 XX Isolated nucleic acid for, e.g. to halogenate various substrates
 PT including proteins, has polynucleotide sequence encoding vanadium
 PT haloperoxidase polypeptide with catalytic helical frame.
 XX
 PS Claim 15; Page 53-56; 56pp; English.
 XX
 CC The invention relates to nucleic acid encoding vanadium haloperoxidase
 CC particularly vanadium bromoperoxidase polypeptide which consists of a
 CC catalytic helical frame that complexes a vanadium ion and catalyses the
 CC oxidation of o-dianisidine (ODA). The invention also provides
 CC recombinantly produced vanadium haloperoxidases. The polypeptides of the
 CC invention can be fused to other proteins to allow quantification or
 CC localisation of the linked protein. They can be used to halogenate
 CC various substrates including proteins; to produce epoxides from alkenes,
 CC halogenated ketones from alkenes; to produce alpha, gamma-halohydrins
 CC from cyclopropanes and to produce dihalogenated products from alkenes and
 CC alkenes. They can also be used in signal generating systems in place of
 CC horseradish peroxidase; as a component in assays; as enzymatic
 CC antimicrobial agents; and in the production of phenolic adhesives. The
 CC polypeptides can be used in industrial catalysis in a variety of
 CC contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
 CC reactions. The present sequence is Fucus distiches vanadium
 CC haloperoxidase. (Updated on 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 676 AA;

Query Match 100.0%; Score 1039; DB 5; Length 676;
 Best Local Similarity 100.0%; Pred. No. 4.6e-114;
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VNFSTSHYFRLIGAAELAQASCYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 60
 DB 435 VNFSTSHYFRLIGAAELAQASCYQKQVHRFARPEALGGLTHTNTAGDLDADFDISLLE 494
 QY 61 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQNGAFATVVKALIGL 120
 DB 495 NDELLKRVAEINAAQNPNNEVTYLLPQAIQVGSPTSPSGHATQNGAFATVVKALIGL 554
 QY 121 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQWLGHIHYFDPGQGL 180
 DB 555 DRGECFPNVPFSDGDLINFEAGCLTYEGEINKLVANVAFGRQWLGHIHYFDPGQGL 614
 QY 181 LLGETITVRTLHOELMTF 198
 DB 615 LLGETITVRTLHOELMTF 632

RESULT 3
 AAE05750
 ID AAE05750 standard; protein; 598 AA.
 XX AC AAE05750;
 XX DT 24-SEP-2001 (first entry)
 XX DE Corallina officinalis vanadium bromoperoxidase (cVBPO).
 XX KW Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
 KW signal generation; analytical system; anti-microbial agent;
 KW industrial application; biological material; avidin; metabolite;
 KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
 KW vitamin; intact cell; drug; narcotic.
 XX OS Corallina officinalis.
 XX FH Key Location/Qualifiers
 FT Active-site 400
 FT /note= "Site at which vanadate cofactor binds"
 FT Active-site 408
 FT /note= "Site at which vanadate cofactor binds"
 FT Misc-difference 474
 FT /note= "Ala is present at this location in the sequence
 FT shown in sequence listing and figure 2 of the
 FT specification"
 FT Active-site 547
 FT /note= "Site at which vanadate cofactor binds"
 FT Active-site 553
 FT /note= "Site at which vanadate cofactor binds"
 FT WO200153494-A2.
 XX PN 26-JUL-2001.
 XX PD 24-JAN-2001; 2001WO-US002383.
 XX PF 24-JAN-2000; 2000US-0177764P.
 XX PR (REGC) UNIV CALIFORNIA.
 XX PA Butler A, Carter JN;
 XX PI WPI; 2001-457611/49.
 XX DR N-PSDB; AAD11050.
 XX PT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
 PT obtained from Corallina officinalis for use in signal generation in
 PT analytical systems.
 XX PS Claim 9; Fig 1; 68pp; English.
 XX CC The present sequence is Corallina officinalis (marine red algae) vanadium
 CC bromoperoxidase (cVBPO). The vanadium haloperoxidase, especially vanadium
 CC bromoperoxidase (VBPO) is useful for signal generation in analytical
 CC systems, and as anti-microbial agents. The VBPO is also useful for
 CC industrial applications. The VBPO is also useful for detecting a wide
 CC variety of chemical and biological materials such as amino acids,
 CC peptides, polypeptides, proteins including enzymes, avidin, antibodies
 CC and antigenic proteins, carbohydrates including monosaccharides,
 CC polysaccharides and lipopolysaccharides, hormones such as human chorionic
 CC gonadotropin, thyroxine, follicle stimulating hormone, luteinising hormone,
 CC hormone, metabolites such as glucose, lactate and pyruvate,
 CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
 CC cells from various organisms including microorganisms and drugs such as
 CC narcotics, therapeutic and those abused
 XX SQ Sequence 598 AA;
 Query Match 32.0%; Score 332.5; DB 4; Length 598;
 Best Local Similarity 38.5%; Pred. No. 5.7e-30;

Matches 82; Conservative 38; Mismatches 72; Indels 21; Gaps 7;
 Qy 1 VNFCTSHYRLI-GAAELAQASCYQKWVHRFARPEALGGTLH-NTIAGD----LDADF 54
 Db 374 VNFQDAHVLSLVTETVATKALKAVRYQKFNHRRRLRPEATGGLISVKNIAAEKGSVPPEV 433
 Qy 55 DISLENDELKRVAEINAAQN-----PNNEVTYLLPQATQVGSPTTPSPSGHATONGA 109
 Db 434 DLAVEELEDILKAEISNRKQNIADGDPDPSPFLPQAFAGSPFPHPSYSGSHAVVAGA 493
 Qy 110 FATVLKALIG---LDRGGCFNPVPSPDDGLELINFEGACLTVEGEINKLAVNVAFGR 165
 Db 494 CVTLKAFPSNQIDQVFE-----VDKDEKLWKSFKGT-LTVAGELNKLADNIAGR 547
 Qy 166 QMLGIHYRFDGIGLLGETITVTRLHQELMTF 198
 Db 548 NMAGVHYPSDQFESILLGEQVAIGILEEQSLTY 580
 RESULT 4
 AAE05756
 ID AAE05756 standard; protein; 598 AA.
 XX AC AAE05756;
 XX DT 24-SEP-2001 (first entry)
 XX DE Corallina pilulifera vanadium bromoperoxidase 1 (VBPO1).
 XX KW Vanadium bromoperoxidase; VBPO; haloperoxidase; signal generation;
 KW analytical system; anti-microbial agent; industrial application;
 KW biological material; avidin; metabolite; lipopolysaccharide;
 KW human chorionic gonadotropin; luteinising hormone; vitamin; intact cell;
 KW drug; narcotic.
 XX OS Corallina pilulifera.
 XX PN WO200153494-A2.
 XX PD 26-JUL-2001.
 XX PF 24-JAN-2001; 2001WO-US002383.
 XX PR 24-JAN-2000; 2000US-0177764P.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI Butler A, Carter JN;
 XX DR WPI; 2001-457611/49.
 XX PT Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
 PT obtained from Corallina officinalis for use in signal generation in
 PT analytical systems.
 XX PS Example 1; Fig 2; 68pp; English.
 XX CC The present sequence is Corallina pilulifera vanadium bromoperoxidase 1
 CC (VBPO1). The vanadium haloperoxidase, especially vanadium bromoperoxidase
 CC (VBPO) is useful for signal generation in analytical systems, and as anti-
 CC microbial agents. The VBPO is also useful for industrial applications.
 CC The VBPO is also useful for detecting a wide variety of chemical and
 CC biological materials such as amino acids, peptides, polypeptides,
 CC proteins including enzymes, avidin, antibodies and antigenic proteins,
 CC carbohydrates including monosaccharides, polysaccharides and
 CC lipopolysaccharides, hormones such as human chorionic gonadotropin,
 CC thyroid stimulating hormone, luteinising hormone, thyroxine, follicle
 CC stimulating hormone, parathyroid hormone and growth hormone, metabolites
 CC such as glucose, lactate and pyruvate, oligonucleotides, nucleic acids,
 CC vitamins such as B12 and biotin, intact cells from various organisms
 CC including microorganisms and drugs such as narcotics, therapeutic and
 CC those abused
 XX

```
SQ Sequence 598 AA;
Query Match 31.9%; Score 331.5; DB 4; Length 598;
Best Local Similarity 40.1%; Pred. No. 7.5e-30;
Matches 85; Conservative 34; Mismatches 74; Indels 19; Gaps 8;

Qy 1 VNFGTSHYFRLI-GAAELAQASCYQKWQVHRFARPEALGGTTLH-NTIA-----GDLDAF 54
Db 374 VNFQSAHVLSLVTVEATKALKAVRYQKFNHRRLRPEATGGLISVKNIAQKGESIFPEV 433

Qy 55 DISLENDELLKRVAEINAAQN-----PNNEVTYLLPOAIQVGSPTSPSHGATONGA 109
Db 434 DLAVEELGDILEKAEISNRKQNIADGDDPDPSPFLPMFAEGSPFPHSYGSHAVVAGA 493

Qy 110 FATVLKALIGDRGCEFPNPVPFSD---DGLLEINFEAGCLTYEGEINKLVNVAFRQ 166
Db 494 CVTILKAF--FDSGIEI--DQVFEVDKDEDKLVKSSFKGT-LTVAGELNKLADNIAIGRN 548

Qy 167 MLGIHYRFDGIQGLLGETITVTRLHQELMTF 198
Db 549 MAGVHYFSDQFESLLGGEQVAIGILEQSLTY 580

RESULT 5
AAW80550
ID AAW80550 standard; protein; 598 AA.
XX
AC AAW80550;
XX
XX 09-DEC-1998 (first entry)
XX
XX Haloperoxidase enzyme, CP.BP01.
XX
XX Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
XX bromine; iodine.
XX
XX Corallina pilulifera.
XX
XX JPI0248581-A.
XX
XX 22-SEP-1998.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX 06-MAR-1997; 97JP-00070539.
XX
XX (SAKA ) OTSUKA PHARM CO LTD.
XX
XX WPI; 1998-560733/48.
XX
XX N-PSDB; AAV56020.
XX
XX New haloperoxidase gene - and corresponding vector, transformed host cell
XX and method of preparation.
XX
XX Claim 1; Page 10-11; 20pp; Japanese.
XX
XX This represents a haloperoxidase enzyme. CP.BP01. The haloperoxidase
XX genes encoding the enzymes CP.BP01 and CP.BP02 are isolated from
XX Corallina pilulifera. A host cell transformed with a vector containing
XX the haloperoxidase genes can be used for the recombinant production of
XX the enzyme. The haloperoxidases of the invention are vanadium-dependent
XX and have halide specificity of bromine and iodine
XX

SQ Sequence 598 AA;
Query Match 31.7%; Score 329.5; DB 2; Length 598;
Best Local Similarity 40.1%; Pred. No. 1.3e-29;
Matches 85; Conservative 34; Mismatches 74; Indels 19; Gaps 8;

Qy 1 VNFGTSHYFRLI-GAAELAQASCYQKWQVHRFARPEALGGTTLHNTIA-----GDLDAF 54
Db 374 VNFQSAHVLSLVTVEATKALKAVRYQKFNHRRLRPEATGGLIIVNIAQKGESIFPEV 433

Qy 55 DISLENDELLKRVAEINAAQN-----PNNEVTYLLPOAIQVGSPTSPSHGATONGA 109
Db 434 DLAVEELGDILEKAEISNRKQNIADGDDPDPSPFLPMFAEGSPFPHSYGSHAVVAGA 493

Qy 110 FATVLKALIGDRGCEFPNPVPFSD---DGLLEINFEAGCLTYEGEINKLVNVAFRQ 166
Db 494 CVTILKAF--FDSGIEI--DQVFEVDKDEDKLVKSSFKGT-LTVAGELNKLADNIAIGRN 548

Qy 167 MLGIHYRFDGIQGLLGETITVTRLHQELMTF 198
Db 549 MAGVHYFSDQFESLLGGEQVAIGILEQSLTY 580

RESULT 6
AAE05755
ID AAE05755 standard; protein; 598 AA.
XX
AC AAE05755;
XX
XX 24-SEP-2001 (first entry)
XX
XX Corallina officinalis vanadium bromoperoxidase mutant (H480A).
XX
XX Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
XX signal generation; analytical system; anti-microbial agent;
XX industrial application; biological material; avidin; metabolite;
XX lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
XX vitamin; intact cell; drug; narcotic; mutant; variant.
XX
XX Corallina officinalis.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 480
XX /note= "wild type His is substituted with Ala"
XX
XX WO200153494-A2.
XX
XX 26-JUL-2001.
XX
XX 24-JAN-2001; 2001WO-US0002383.
XX
XX 24-JAN-2000; 2000US-0177764P.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Butler A, Carter JN;
XX
XX WPI; 2001-457611/49.
XX
XX Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
XX obtained from Corallina officinalis for use in signal generation in
XX analytical systems.
XX
XX Example 5; Page; 68pp; English.
XX
XX The present sequence is Corallina officinalis (marine red algae) vanadium
XX bromoperoxidase (cVBPO) mutant (H480A). The vanadium haloperoxidase,
XX especially vanadium bromoperoxidase (VBPO) is useful for signal
XX generation in analytical systems, and as anti-microbial agents. The VBPO
XX is also useful for industrial applications. The VBPO is also useful for
XX detecting a wide variety of chemical and biological materials such as
XX amino acids, peptides, polypeptides, proteins including enzymes, avidin,
XX antibodies and antigenic proteins, carbohydrates including
XX monosaccharides, polysaccharides and lipopolysaccharides, hormones such
XX as human chorionic gonadotropin, thyroid stimulating hormone, luteinising
XX hormone, thyroxine, follicle stimulating hormone, parathyroid hormone and
XX growth hormone, metabolites such as glucose, lactate and pyruvate,
XX oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
XX cells from various organisms including microorganisms and drugs such as
XX narcotics, therapeutic and those abused. Note: The present sequence is
XX not shown in the specification but is derived from the wild type
XX Corallina officinalis vanadium bromoperoxidase (cVBPO) sequence shown in
```

Sequence 598 AA;

Query Match 31.0%; Score 322.5; DB 4; Length 598;

Best Local Similarity 38.0%; Pred. No. 8.8e-29;
Matches 81: Conservative 38: Mismatches 73: Indels 21: Gaps 7:

1 INEECTSHVERB.I -CAAEI AOBASCYOWUWUPEAPDEAYCCTIY-NNTIACD - - - I DANE 54

[illegible]

1

35 DISLTENDELLKRRVAEINAAQN-----FNNEVTYLLLPQATQVGSPTHPSYPSPGHATQNGA 109

434 DLAVEELEDILEKAEISNRKQNIADGDPDPDPFLLPQAFAGSPFAPSYGSGHAVVAGA 493

110 FATVLKALIG---LDRGGECPNPVFPSPDDGLELINFEGACLTYEGEINKLAVNVAFR 165

494 C V T I L K A F F D S N F Q I D Q V F E - - - - - V D K D E D K L V K S S F K G T - L T V A G E L N K L A D N I A I G R 547

166 OMLGIHYRFDGIOGLLGETITVRTLHOELMTF 198

549 NMACHVESNOEESTILICEOVATICTIPEASITY 590

RESULT 7

AAW80551 standard: protein: 597 AA.

X 1700551 .

09-DEC-1998 (TISC entry)

Haloperoxidase enzyme, CP.BP02.

Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;

OUTSIDE (CONTINUED)

СОСТАВНА ПРИЛИКА.

JP10248581-A.

22-SEP-1998.

06-MAR-1997; 97JP-00070539.

06-MAR-1997; 97JP-00070539.

(SAKA) OTSUKA PHARM CO LTD.

0
1
2
3
4
5
6
7
8
9

N-PSDB; AAV56021.

New haloperoxidase gene - and corresponding vector, transformed host cell

and method of preparation.

Claim 1; Page 11-13; 20pp; Japanese.

This represents a haloperoxidase enzyme, CP.BP02. The haloperoxidase

genes encoding the enzymes CP.BP01 and CP.BP02 are isolated from *Corallina pilulifera*. A host cell transformed with a vector containing

the haloperoxidase genes can be used for the recombinant production of the enzyme. The haloperoxidases of the invention are vanadium-dependent:

and have halide specificity of bromine and iodine

Sequence 597 AA;

Query Match

Best Local Similarity 39.7%; Pred. No. 1e-28;
Matches 87. Conservative 39. Mismatches 71. Indels 32. Cans 8.

[illegible]

	Matches	87;	Conservative	29;	Mismatches	71;	Indels	32;	Gaps	8;
Qy	1	VNFCTSYFRLI	-GAAELAQRASCYQKWVHRFARPEALGGTL-----HNTIAGDLDAFD	55						
		:	:	:	:	:	:	:	:	:
Db	373	VNFGSAHVLSLVTRVATRALKAVRYKFNIHRRLRPEATGGII	SVNNKSFLAQS----	427						
		:	:	:	:	:	:	:	:	:
Qy	56	ISLLENDELKR-----VAEINAAQNPNNEV-----TYLLPQAIQVGSPTHPSYSGCH	103							
		:	:	:	:	:	:	:	:	:
Db	428	IIPFEVSELVEELSSIIILDDVAESNEKQNRADGIIVDPDKSFLLPWAFAGSPFPHPYSYGSH	487							
		:	:	:	:	:	:	:	:	:
Qy	104	ATQNGAFATVLKALIG----LDRCGECPNPVPFSDDGLELINFEAGCLTVEGINKLAV	159							
		:	:	:	:	:	:	:	:	:
Db	488	AVVAGACVTILKAFFDFANFOIDKYFE-----VDTDDEKLVKSSPKGT-LTVAGELINKLAD	541							
		:	:	:	:	:	:	:	:	:
Qy	160	NVAFGROMLGHYRFDGIQGILLIGETTITVRTLHQELMTFF	198							
		:	:	:	:	:	:	:	:	:
Db	542	NVAIGRNMGVHYPSDFOPESLLLIGEQAIGLIEEQSLITY	580							
		:	:	:	:	:	:	:	:	:

RESULT 9	
AAE17502	
ID	AAE17502 standard; protein; 22 AA.
XX	
XX	AAE17502;
XX	
DT	07-AUG-2003 (revised)
DT	01-MAY-2002 (first entry)
XX	
DE	Fucus distiches vanadium haloperoxidase vanadium-binding region #1.
DE	
KW	Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
KW	epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
KW	industrial catalysis; enzyme; catalyst.

```

CC  to correct OS field.)
XX
SQ  Sequence 22 AA;

Query Match      11.7%; Score 122; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 4e-07;
Matches 22; Conservative. 0; Mismatches 0; Indels 0; Gaps 0;

QY  18 AGRASCYQKQVHRFARPEALG 39
DB  1 AGRASCYQKQVHRFARPEALG 22
    |||||
RESULT 10
AAEI7504
ID  AAEI7504 standard; protein; 21 AA.
XX  AAEI7504;
XX
XX  07-AUG-2003 (revised)
DT  01-MAY-2002 (first entry)
XX
XX  Fucus distiches vanadium haloperoxidase vanadium-binding region #3.
DE  Vanadium haloperoxidase; vanadium bromoperoxidase; o-dianisidine; ODA;
KW  epoxide; antimicrobial agent; phenolic adhesive; halohydrin;
KW  industrial catalysis; enzyme; catalyst.
XX  Fucus distichus.
OS  OS
XX  WO200200838-A2.
XX
XX  03-JAN-2002.
XX
XX  07-JUN-2001; 2001WO-US018602.
XX
XX  19-JUN-2000; 2000US-00596794.
XX
XX  (REGC ) UNIV CALIFORNIA.
PA
XX  Vreeland V;
PI
XX  WPI; 2002-154651/20.
DR
XX
XX  Isolated nucleic acid for, e.g. to halogenate various substrates
PT  including proteins, has polynucleotide sequence encoding vanadium
PT  haloperoxidase polypeptide with catalytic helical frame.
XX
PS  Example 1; Fig 5; 56pp; English.
XX
XX  The invention relates to nucleic acid encoding vanadium haloperoxidase
CC  particularly vanadium bromoperoxidase polypeptide which consists of a
CC  catalytic helical frame that complexes a vanadium ion and catalyses the
CC  oxidation of o-dianisidine (ODA). The invention also provides
CC  recombinantly produced vanadium haloperoxidases. The polypeptides of the
CC  invention can be fused to other proteins to allow quantification or
CC  localisation of the linked protein. They can be used to halogenate
CC  various substrates including proteins; to produce epoxides from alkenes,
CC  halogenated ketones from alkynes; to produce alpha, gamma-halohydrins
CC  from cyclopropanes and to produce dihalogenated products from alkenes and
CC  alkynes. They can also be used in signal generating systems in place of
CC  horseradish peroxidase; as a component in assays; as enzymatic
CC  antimicrobial agents; and in the production of phenolic adhesives. The
CC  polypeptides can be used in industrial catalysis in a variety of
CC  contexts, e.g. catalyst for halogenation, oxidation, and epoxidation
CC  reactions. The present sequence is Fucus distiches vanadium
CC  haloperoxidase conserved vanadium-binding region. (Updated on 07-AUG-2003
CC  to correct OS field.)
XX
SQ  Sequence 21 AA;

Query Match      10.6%; Score 110; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 1e-05;

```

Qy 94 PTHPSYPGSHATQNGAFAT 112
Db 1 PTHPSYPGSHATQNGAFAT 19

Db 137 DVAVGVALGATV 148

RESULT 13
AAE05754
ID AAE05754 standard; peptide; 37 AA.
XX AC AAE05754;
XX
DT 24-SEP-2001 (first entry)
XX
DE Vanadium bromoperoxidase (VBPO) consensus sequence #2.
XX
XX Marine red algae; vanadium bromoperoxidase; VBPO; haloperoxidase;
KW signal generation; analytical system; anti-microbial agent;
KW industrial application; biological material; avidin; metabolite;
KW lipopolysaccharide; human chorionic gonadotropin; luteinising hormone;
KW vitamin; intact cell; drug; narcotic.
XX
OS Corallina officinalis.
OS Ascophyllum nodosum.
OS Fucus distichus.
XX
XX Key Location/Qualifiers
FH Misc-difference 3 /label= Unknown
FT Misc-difference 4 /label= Unknown
FT Misc-difference 7 /label= Unknown
FT Misc-difference 12 /label= Ile, Leu
FT Misc-difference 14 /label= Unknown
FT Misc-difference 16 /label= Unknown
FT Misc-difference 19 /label= Unknown
FT Misc-difference 21 /label= Unknown
FT Misc-difference 23 /label= Unknown
FT Misc-difference 26 /label= Unknown
FT Misc-difference 27 /label= Unknown
FT Misc-difference 29 .32 /label= Unknown
XX
XX WO200153494-A2.
XX
XX 26-JUL-2001.
XX
XX 24-JAN-2001; 2001WO-US002383.
XX
XX 24-JAN-2000; 2000US-0177764P.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Butler A, Carter JN;
XX
XX WPI; 2001-457611/49.
XX
XX Novel vanadium haloperoxidase, especially vanadium bromoperoxidase,
PT obtained from Corallina officinalis for use in signal generation in
PT analytical systems.
XX
XX Example 1; Page 41; 68pp; English.
XX
XX The present sequence is a vanadium bromoperoxidase (VBPO) consensus
CC sequence which is derived from Corallina officinalis, Corallina
CC pilulifera, Ascophyllum nodosum and Fucus distichus. The vanadium
CC haloperoxidase, especially vanadium bromoperoxidase (VBPO) is useful for
CC
CC signal generation in analytical systems, and as anti-microbial agents.
CC The VBPO is also useful for industrial applications. The VBPO is also
CC useful for detecting a wide variety of chemical and biological materials
CC such as amino acids, peptides, polypeptides, proteins including enzymes,
CC avidin, antibodies and antigenic proteins, carbohydrates including
CC monosaccharides, polysaccharides and lipopolysaccharides, hormones such
CC as human chorionic gonadotropin, thyroid stimulating hormone, luteinising
CC hormone, thyroxine, follicle stimulating hormone, parathyroid hormone and
CC growth hormone, metabolites such as glucose, lactate and pyruvate.
CC oligonucleotides, nucleic acids, vitamins such as B12 and biotin, intact
CC cells from various organisms including microorganisms and drugs such as
CC narcotics, therapeutic and those abused
XX
SQ Sequence 37 AA;
Query Match 9.6%; Score 100; DB 4; Length 37;
Best Local Similarity 59.5%; Pred. No. 0.00037;
Matches 22; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 148 LTYEGEINKLVNVAFGROMLGTHYRFDGIGQLLGE 184
DB 1 LTYXGXNKLVNVAFGROMLGTHYRFDGIGQLLGE 37
RESULT 14
ABB70948
ID ABB70948 standard; protein; 396 AA.
XX AC ABB70948;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39636.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL15051.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 39636; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ffp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 396 AA;

This Page Blank (uspio)

A;Accession: C70888
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-165 <COL>

Qy	55	DISLENDELKRVAEINAAQNPNNEVYLLPQAIQVSGSPHSPYSPSGHATONGAFATVL	114
Db	64	-----VLIKRLVR---RQPDHPAI-----AVNVDTPSQLSPSAHATSTTAAALLM	107
Qy	115	KALIGLDRGGECFNPVPFPPSDGLELINFEGACTYEIGEINKLAVNVAFGQMLGIHYRF	174
Db	108	GRATGL-----PLPVV-----LVPPMALSRILGLGVHYP	136
Qy	175	DGIQGLLLGETI	186
Db	137	DVAVGVALGATV	148

	Query Match	8.3%	Score 86;	DB 2;	Length 192;
	Best Local Similarity	24.4%;	Pred. No. 0.96;		
	Matches 31;	Conservative 16;	Mismatches 44;	Indels 36;	Gaps 4;
Qy	64	LLKRVAEINAAQNPNNEVTVLLPQATQVCSPTHPSPGSHATQNGAFATVVLKALIGLDRG	123		
Db	78	LKKRWV---RIKPNHAV-----TVNVTGTPSLSPFSAHATSTAAAILIGRASRLPKG	129		
Qy	124	GECPNPVPSPDDGLINFEAGCLTYEINKLVANVAFGRQMLGHIHVRFDIGQGLLIG	193		
Db	130	-----IVAAV-----LVAPMALSRVLGVHVPSPDVAFGVVLG	161		

```

Cy      184 ETITVRT 190
      :
Db      162 AAVAGTT 168

RESULT 6
S69334
chloride peroxidase (EC 1.11.1.10) [validated] - Curvularia inaequalis
N:Alternate names: vanadium-containing chloroperoxidase
C:Species: Curvularia inaequalis
C:Date: 28-Jan-2000 #sequence_revision 28-Jan-2000 #text_change 15-Sep-2000
C:Accession: S69334; S53117

```


R;Simons, B.H.; Barnett, P.; Vollenbroek, E.G.M.; Dekker, H.L.; Muijers, A.O.; Messers
Eur. J. Biochem. 229, 566-574, 1995
A;Title: Primary structure and characterization of the vanadium chloroperoxidase from th
A;Reference number: S69334; MUID:95262722; PMID:7744081
A;Accession: S69334
A;Molecule type: mRNA
A;Residues: 1-609 <SIM>
A;Cross-references: EMBL:X85369; NID:g732617; PIDN:CAA59686.1; PID:g732618
A;Experimental source: CBS 102.42
A;Note: Part of this sequence was confirmed by sequence analysis of genomic DNA
R;Messerschmidt, A.; Weyer, R.
submitted to the Brookhaven Protein Data Bank, September 1995
A;Reference number: A66871; PDB:1VNC
A;Contents: annotation; X-ray crystallography, 2.1 angstroms, 3-543, 'E', 545-578
A;Note: the authors believe that 544-Glu is correct from modeling evidence
C;Genetics:
A;Gene: vCPO
C;Function:
A;Description: catalyzes the formation of two carbon-chlorine bonds and two water molecu
C;Superfamily: Curvularia inaequalis chloride peroxidase; glucose-6-phosphatase catalyti
C;Keywords: blocked amino end; chloride; metalloprotein; oxidoreductase; vanadium
F;329-511/Domain: Glucose-6-phosphatase catalytic domain homology <GPH>
F;404/Active site: His #status predicted
F;496/Binding site: vanadate (His) #status predicted

Query Match 8.2%; Score 85.5; DB 1; Length 609;
Best Local Similarity 21.7%; Pred. No. 5;
Matches 30; Conservative 22; Mismatches 35; Indels 51; Gaps 5;
QY 71 INAAQNPNNEVTLLPQAIQVSGTPSPYSGHATQNGAFATVLK-----115
Db 381 LGPATNTNDIPF-----RPPFPAYPSGHATFGGAVQVMRRYNGRVGTWKDDEP 431
QY 116 -----ALIGDRGCECPNPVPSPDDGLELI-----NFEGAC-LTVEGEINKL 157
Db 432 DNTAIDMISEELNGVNRDLRQYDPTAPIEDQGVTRIVRHFPDSAWELMFENALISRI 491
QY 158 ANNVAFGRQMLGIHYRFD 175
Db 492 -----FLGVHWRFD 500

RESULT 7
D97672
hypothetical protein AGR_C_4732 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 02-Jun-2003
C;Accession: D97672
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D97672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88333.1; PID:g15157811; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4732
A;Map position: circular chromosome
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.1%; Score 84.5; DB 2; Length 217;
Best Local Similarity 29.3%; Pred. No. 1.6;
Matches 27; Conservative 8; Mismatches 28; Indels 29; Gaps 2;
QY 92 GSPTSPYSGHATQNGAFATVLKALIGDRGCECPNPVPSPDDGLELINFEGACLTYE 151
Db 116 GDNLYESFPFGHSTAGAFPGVFAML-----MRFERWAFLL-- 151

QY 152 GEINKLVANVAFGRQMLGIHYRFDGIGLLG 183
Db 152 GEINKLVANVAFGRQMLGIHYRFDGIGLLG 183

Db 152 -----LALVIGSRVIVGAHYHPSDVAAGLLLG 178

RESULT 8
AB2897
conserved hypothetical protein Atu2611 [imported] - Agrobacterium tumefaciens (strain C58
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 02-Jun-2003
A;Accession: AB2897
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB2897
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <KUR>
A;Cross-references: GB:AE008698; PIDN:AAI43592.1; PID:g17741109; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2611
A;Map position: circular chromosome
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom

Query Match 8.1%; Score 84.5; DB 2; Length 242;
Best Local Similarity 29.3%; Pred. No. 1.8;
Matches 27; Conservative 8; Mismatches 28; Indels 29; Gaps 2;
QY 92 GSPTSPYSGHATQNGAFATVLKALIGDRGCECPNPVPSPDDGLELINFEGACLTYE 151
Db 141 GDNLYESFPFGHSTAGAFPGVFAML-----MRFERWAFLL-- 176

QY 152 GEINKLVANVAFGRQMLGIHYRFDGIGLLG 183
Db 177 -----LALVIGSRVIVGAHYHPSDVAAGLLLG 203

RESULT 9
AF2306
hypothetical protein alr4005 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C;Accession: AF2306
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.;
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2306
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-418 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA075704.1; PID:g17133140; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4005

Query Match 8.1%; Score 84; DB 2; Length 418;
Best Local Similarity 20.8%; Pred. No. 4.2;
Matches 49; Conservative 29; Mismatches 93; Indels 66; Gaps 9;
QY 4 GTSHYFRLIGAAELARASCYKQW-----VHRFA----- 33
Db 98 GGDQFFVVGKLGKGYRTVYAEWEARWHNLIDRFGVMKPLVAAKASPKYAHKFTVVVDL 157

QY 34 -----RPEALGGLTHTIAGDLADFDLSLENEDELLKRVAEINAQNPNEVTYLLP 86
Db 158 MLEANAQLPNAPHTPTHTPIIGILPGSKAAKLTGGIPLMLSIGEYIHSKRP--ETKEVIP 215

Query Match	7.7%; Score 80; DB 2; Length 489;	
Best Local Similarity	34.9%; Pred. No. 12;	
Matches	30; Conservative	9; Mismatches 39; Indels 8; Gaps 4;
Qy	26 KQVHREARPEALGGLTILNTIA-GDLDADEISLLENDELLKRVAEINAAQNPNNEVYL 84	
Db	290 KQOHRLLVNEKLTQTHRIYACGDVTCGYDLPVANYE--ARTALNNALFLRLQVNYQ 347	
Qy	85 -LPOAIQVGSPTHPSPYSGHATONGA 109	
Db	348 NTPWAIL-----SHPTLAQVGVTEHQ 369	
RESULT 14		
A64822		
probable bacitracin transport permease ybjG - Escherichia coli (strain K-12)		
C;Species: Escherichia coli		
C;Date: 28-Jan-2000 #sequence_revision 28-Jan-2000 #text_change 01-Mar-2002		
C;Accession: A64822		
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co		
.A.; Rose, D.J.; Mau, B.; Shao, Y.		
Science 277, 1453-1462, 1997		
A;Title: The complete genome sequence of Escherichia coli K-12.		
A;Reference number: A64720; MUID:97426617; PMID:9278503		
A;Accession: A64822		
A;Status: nucleic acid sequence not shown; translation not shown		
A;Molecule type: DNA		
A;Residues: 1-198 <BLAT>		
A;Cross-references: GB:AE000186; GB:U00096; NID:gl787058; PIDN:AACT3928.1; PID:gl787064;		
A;Experimental source: strain K-12, substrain MG1655		
C;Genetics:		
A;Gene: ybjG		
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom		
C;Keywords: antibiotic resistance; transmembrane protein		
F;34-50/Domain: transmembrane #status predicted <TM1>		
F;51-160/Domain: glucose-6-phosphatase catalytic domain homology <GPH>		
F;56-72/Domain: transmembrane #status predicted <TM2>		
F;128-144/Domain: transmembrane #status predicted <TM3>		
F;151-167/Domain: transmembrane #status predicted <TM4>		
Query Match	7.7%; Score 79.5; DB 1; Length 198;	
Best Local Similarity	36.8%; Pred. No. 4.2;	
Matches	21; Conservative	8; Mismatches 23; Indels 5; Gaps 1;
Qy	132 FPSDDGLELINFEGACLTYE-----GEINKLVNVAFGQMLGIHYRFDGIQGLLLG 183	
Db	101 FPSDHGTVITTFALAFLCWHLWSGSLMLVAVVIAWSRVYLGVHWPLDMLGGLLAG 157	
RESULT 15		
A99744		
hypothetical protein ECs0921 [imported] - Escherichia coli (strain O157:H7, substrain R		
C;Species: Escherichia coli		
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001		
C;Accession: A99744		
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.		
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.		
DNA Res. 8, 11-22, 2001		
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend		
A;Reference number: A99629; MUID:21156231; PMID:11258796		
A;Accession: A99744		
A;Status: preliminary		
A;Molecule type: DNA		
A;Residues: 1-198 <HAY>		
A;Cross-references: GB:BA000007; PIDN:BA034344.1; PID:gl3360380; GSPDB:GN00154		
A;Experimental source: strain O157:H7, substrain RIMD 0509952		
C;Genetics:		
A;Gene: ECs0921		
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom		
Query Match	7.7%; Score 79.5; DB 2; Length 198;	
Best Local Similarity	36.8%; Pred. No. 4.2;	

Matches	21; Conservative	8; Mismatches 23; Indels 5; Gaps 1;
Qy	132 FPSDDGLELINFEGACLTYE-----GEINKLVNVAFGQMLGIHYRFDGIQGLLLG 183	
Db	101 FPSDHGTVITTFALAFLCWHLWSGSLMLVAVVIAWSRVYLGVHWPLDMLGGLLAG 157	
Search completed:	September 19, 2004, 03:59:49	
Job time :	13.0069 secs	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 03:16:07 ; Search time 36.7002 Seconds
(without alignments)
1702.240 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632

Perfect score: 1039

Sequence: 1 VNFQSHYFRLIGAELAQR.....GLLGGTITVTLHQELMTF 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	100.0	676	10	O82433
2	503.5	48.5	682	10	Q7X9V0
3	502.5	48.4	646	10	Q7X9V1
4	332.5	32.0	598	10	Q8LLW7
5	331.5	31.9	598	10	O81959
6	322	31.0	597	10	O81960
7	116.5	11.2	145	16	Q9RVL3
8	108	10.4	165	16	O53584
9	104	10.4	165	16	Q7TVM5
10	104	10.0	416	16	Q89ER7
11	99	9.5	706	16	Q7UUV2
12	93	9.0	525	16	Q825G8
13	92.5	8.9	299	10	Q94AY8
14	90.5	8.7	441	10	Q8L738
15	90.5	8.7	667	13	Q7T2E5
16	88.5	8.5	396	5	Q9VDN3

17	88.5	8.5	647	4	Q9H6T8
18	87	8.4	198	16	Q8ZGH3
19	87	8.4	201	16	Q8D031
20	87	8.4	308	2	Q9RFH0
21	86.5	8.3	248	3	P78610
22	86.5	8.3	441	10	O23053
23	86.5	8.3	487	2	P72262
24	86	8.3	192	16	Q9CDB6
25	86	8.3	201	17	Q8TW45
26	85.5	8.2	367	11	Q8R2X1
27	85.5	8.2	881	11	Q8C0E7
28	84.5	8.1	242	16	Q8UC87
29	84.5	8.1	289	16	Q7TUQ8
30	84.5	8.1	1431	10	Q7XR16
31	84	8.1	168	16	Q8NLQ8
32	84	8.1	307	16	Q8FLZ6
33	84	8.1	418	16	Q8YQ31
34	84	8.1	453	16	Q82PX4
35	84	8.1	612	16	Q88VB9
36	83.5	8.0	626	4	Q9NXC4
37	83	8.0	271	16	Q98GT9
38	83	8.0	421	16	Q7UHC0
39	83	8.0	465	16	Q89AP5
40	83	8.0	800	13	Q918X3
41	82	7.9	308	2	Q9RFH7
42	82	7.9	308	2	Q9RFH2
43	81.5	7.8	235	16	Q92KY7
44	81.5	7.8	487	2	P72269
45	81.5	7.8	619	2	Q8VTN2

ALIGNMENTS

RESULT 1

O82433 PRELIMINARY; PRT; 676 AA.
 AC O82433;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Vanadium bromoperoxidase.
 OS Fucus distichus.
 OC Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Fucus.
 OX NCBI_TaxID=3012;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Vreeland V., Ng K., Epstein L.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF053411; AAC35279.1; -;
 DR HSP; P81701; I019.
 DR GO; GO:0004601; Fiperoxidase activity; IEA.
 DR InterPro; IPR008934; ACPase_VanPerase.
 DR InterPro; IPR000326; FA_PTPase.
 KW Peroxidase.
 SQ SEQUENCE 676 AA; 73344 MW; AA8A46F539A14EC7 CRC64;

Query Match	100.0%;	Score	1039;	DB	10;	Length	676;
Best Local Similarity	100.0%;	Pred. No.	1.8e-90;				
Matches	198;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	VNFQSHYFRLIGAELAQRASCYQKWVHRFARPEALGGLTHTNTIAGDLADFDISLLE	60				
DB	435	VNFQSHYFRLIGAELAQRASCYQKWVHRFARPEALGGLTHTNTIAGDLADFDISLLE	494				
QY	61	NDELLKRVABINAAQPNNEVTYLLPOAIOVGSPTSPSGHATQNGAFATVLKALIGL	120				
DB	495	NDELLKRVABINAAQPNNEVTYLLPOAIOVGSPTSPSGHATQNGAFATVLKALIGL	554				
QY	121	DRGECFPNPFVDDGCLLEINFEAGCTVEGEINKLVANVAFGRMLGTHYRFDGQGL	180				
DB	555	DRGECFPNPFVDDGCLLEINFEAGCTVEGEINKLVANVAFGRMLGTHYRFDGQGL	614				

```
QY 181 LLGETITVTRLHQLMTF 198
DB 615 LLGETITVTRLHQLMTF 632

RESULT 2
Q7X9V0 PRELIMINARY; PRT; 682 AA.
AC Q7X9V0;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 2.
GN VBPO2.
OS Laminaria digitata.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Laminariaceae;
OC Laminaria.
OX NCBI_TaxID=80365;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sporophyte;
RX MEDLINE=22703831; PubMed=12697758;
RA Colin C., Leblanc C., Wagner E., Delage L., Leize-Wagner E.,
RA van Dorselaer A., Kloareg B., Potin P.;
RT "The Brown Algal Kelp Laminaria digitata Features Distinct
RT Bromoperoxidase and Iodoperoxidase Activities.";
RL J. Biol. Chem. 278:23545-23552(2003).
DR EMBL; AJ491787; CAD37192.1; -.
KW Peroxidase.
SQ SEQUENCE 682 AA; 75125 MW; 98DD4FC45B17E772 CRC64;

Query Match 48.5%; Score 503.5; DB 10; Length 682;
Best Local Similarity 46.7%; Pred. No. 2.8e-39;
Matches 98; Conservative 28; Mismatches 67; Indels 17; Gaps 2;

QY 3 FGTSHYFRLLIGAEALQASCYQKQVHRPAREALGGTLHNTIAGDLADDFDISLLEND 62
DB 386 FGEFHILTAMASASSSTRHAWYAKQVHRMLRPEAYGALVHNTLMRDVITPLPDSILRNT 445

QY 63 ELLKRVAEINAAONPNEVYLLPQAIQVGSPTSPHSGHATONGAFATVLCALIGLDR 122
DB 446 ELLNRVEVHNQRNPDGKTFLLPMAAQQSPHPAYPSGHAINNGAYITALKAFILGYEA 505

QY 123 GGCFPNPVPSPDDGLELINF-----GAC-----LTYGEINKLAVNVAFGR 165
DB 506 GQKCFPNVVSNDGKTKRIKYPGSGREIVGECVNEKGKLVGELYEGELNKISANVLLGR 565

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQL 195
DB 566 SHIGVHRMDGVYGALMGETSCVRRLOQEL 595

RESULT 3
Q7X9V1 PRELIMINARY; PRT; 646 AA.
AC Q7X9V1;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 1.
GN VBPO1.
OS Laminaria digitata.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales; Laminariaceae;
OC Laminaria.
OX NCBI_TaxID=80365;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sporophyte;
RX MEDLINE=22703831; PubMed=12697758;
RA Colin C., Leblanc C., Wagner E., Delage L., Leize-Wagner E.,
RA van Dorselaer A., Kloareg B., Potin P.;
RT "The Brown Algal Kelp Laminaria digitata Features Distinct
RT Bromoperoxidase and Iodoperoxidase Activities.";
```

```
RL J. Biol. Chem. 278:23545-23552(2003).
DR EMBL; AJ491786; CAD37191.1; -.
KW Peroxidase.
SQ SEQUENCE 646 AA; 71171 MW; DB9382F74CBB97FD CRC64;

Query Match 48.4%; Score 502.5; DB 10; Length 646;
Best Local Similarity 46.7%; Pred. No. 3.2e-39;
Matches 98; Conservative 27; Mismatches 68; Indels 17; Gaps 2;

QY 3 FGTSHYFRLLIGAEALQASCYQKQVHRPAREALGGTLHNTIAGDLADDFDISLLEND 62
DB 350 FGEFHILTAMASASSSTRHAWYAKQVHRMLRPEAYGALVHNTLMRDVITPLPDSILRNT 409

QY 63 ELLKRVAEINAAONPNEVYLLPQAIQVGSPTSPHSGHATONGAFATVLCALIGLDR 122
DB 410 ELLNRVEVHNQRNPDGKTFLLPMAAQQSPHPAYPSGHAINNGAYITALKAFILGYEA 469

QY 123 GGCFPNPVPSPDDGLELINF-----GAC-----LTYGEINKLAVNVAFGR 165
DB 470 GQKCFPNVVSNDGKTKRIKYPGSGREIVGECVNEKGKLVGELYEGELNKISANVLLGR 529

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQL 195
DB 530 SHIGVHRMDGVYGALMGETSCVRRLOQEL 559

RESULT 4
Q8LLW7 PRELIMINARY; PRT; 598 AA.
AC Q8LLW7;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase.
OS Corallina officinalis (Coral seaweed).
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=35170;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22117781; PubMed=12121762;
RA Carter J.N., Beatty K.E., Simpson M.T., Butler A.;
RT "Reactivity of recombinant and mutant vanadium bromoperoxidase from
RT the red alga Corallina officinalis.";
RL J. Inorg. Biochem. 91:59-69(2002).
DR EMBL; AF218810; AM46061.1; -.
DR PDB; 1QHB; 05-JUL-00.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR InterPro; IPR008934; ACPase VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW Peroxidase.
SQ SEQUENCE 598 AA; 65458 MW; BC784E370D748F01 CRC64;

Query Match 32.0%; Score 332.5; DB 10; Length 598;
Best Local Similarity 38.5%; Pred. No. 5.1e-23;
Matches 82; Conservative 38; Mismatches 72; Indels 21; Gaps 7;

QY 1 VNFQSHYFRLLIGAEALQASCYQKQVHRPAREALGGTLH-NTIAGD----LDADF 54
DB 374 VNFQDAVLSLVTEVATRALKAVRYQKFNHRLRPEATGGLISVNKIAAEKGESVPEV 433

QY 55 DISLLENDELKRVAEINAAQN-----PNNEVYLLPQAIQVGSPTSPHSGHATONGA 109
DB 434 DLAVEELEDLEKAEISNRKQNTADGDDPDPPDFLLPQAFAGSPHPSPYSGSHAVVAGA 493

QY 110 FATVLKALIG-----LDRGGECFPNVPSPDDGLELINFEGACITYPEGEINKLAVNVAFGR 165
DB 494 CVTILKAFDPSNFQIDQVPE-----VDKDBDKLVKSFKGT-LTVAGELNKLADNIAIGR 547

QY 166 QMLGIHYRFDGIQGLLIGETITVTRLHQLMTF 198
DB 548 NMAGVHYFSQDFESILLGQVATGAILLEESLTY 580
```

```

RESULT 5
O81959 PRELIMINARY; PRT; 598 AA.
ID O81959 AC O81959;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 1.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimonishi M., Kuwamoto S., Inoue H., Wiewer R., Ohshiro T., Izumi Y., Tanabe T.;
RA "Cloning and expression of the gene for a vanadium-dependent RT bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87657; BAA31261.1; -.
DR HSSP; P81701; I019.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase.VanPerase.
DR InterPro; IPR000326; PA_PTPase.
KW peroxidase.
SQ SEQUENCE 598 AA; 65313 MW; 80B81B678B7FACE0 CRC64;

Query Match 31.9%; Score 331.5; DB 10; Length 598;
Best Local Similarity 40.1%; Pred. No. 6.3e-23;
Matches 85; Conservative 34; Mismatches 74; Indels 19; Gaps 8;

QY 1 VNFGTSHYFLRI-GAAELAQRASCYKQVHRFARPEALGGTLH-NTIA----GDLDADF 54
   |||||: : : : |||: : |||: ||| | : ||| : ||| : : ||| : : ||| : :
Db 374 VNFGSAHVLSLVTEVATRALKA VKRYKFNIHRRLRPEATGLISWNKIAPQGESIIFEV 433

QY 55 DISLENDLKLKRVAEINNAQN-----PNNEVTLLPOAIQVGSPTHPSYSGHATONGA 109
   : :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 434 DLAVEUGDLILEKAEINRKNQNIADGDPDPPSFLLPWAFAGSFFPHPSYSGHAVVAGA 493

QY 110 FATVLKALIGLDRGEGCFNPVPFSD---DGLELINFEGACLTYYEGINKLANVNVAFRQ 166
   : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 494 CVTILKAF--PDSGIEI--DOQFEVDKDCKVKSSFKG-TLVAGELNKLADNIAIGN 548

QY 167 MLGIHVRFDGIQGLLIGETTIVRTLHQELMTF 198
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 549 MAGVHYFSDFESILLGEQAIGILERQSLEY 580

RESULT 6
O81960 PRELIMINARY; PRT; 597 AA.
ID O81960 AC O81960;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Vanadium-dependent bromoperoxidase 2.
OS Corallina pilulifera.
OC Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Corallinaceae;
OC Corallinoideae; Corallina.
OX NCBI_TaxID=78447;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98307393; PubMed=9645486;
RA Shimonishi M., Kuwamoto S., Inoue H., Wiewer R., Ohshiro T., Izumi Y., Tanabe T.;
RA "Cloning and expression of the gene for a vanadium-dependent RT bromoperoxidase from a marine macro-alga, Corallina pilulifera.";
RL FEBS Lett. 428:105-110(1998).
DR EMBL; D87658; BAA31262.1; -.
DR HSSP; P81701; I019.
DR GO; GO:0004601; F:peroxidase activity; IEA.

```



```
Db 49 EYNVAR-POSMBDHVQWAPSLPTPPSPVPSGHATVSGAAAEVL-----AQFPP 98
QY 129 NPVFPSPDDGLELINFEGACITVEGEINKLVANVAFGRQMLGIHYRFGIQGLLGETI 186
Db 99 -----LQARQLRRDARDAAFSRVGGIHWGVGDVAGLDVGQRV 136

RESULT 8
OS3584 PRELIMINARY; PRT; 165 AA.
AC OS3584;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein (PAP2 superfamily protein).
GN RV3807C OR MT0026.12 OR MT3914.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekalia E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagals K., Krogh A., McLean L.A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022076; CAAL17871.1; -.
DR EMBL; AE007185; AAK48280.1; -.
DR PIR; C70888; C70888.
DR TIGR; MT3914; -.
DR TubercuList; RV3807c; -.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF01569; PAP2; 1.
DR SMART; SM00014; acidPPc; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 165 AA; 17218 MW; 0E084FA26508A4A1 CRC64;

Query Match 10.4%; Score 108; DB 16; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.025;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

QY 4 GTSHY-----FRLLGAAELAQASCYQKQVHRFARPEALGGTLHNTIAGDLDADF 54
Db 20 GLSHFGHCIGWLILALLGIALPRR---REMLV-----AGAGAFVAHAIA----- 63
QY 55 DISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTSPSGHATONGAPATVL 114
Db 64 -----VLIKLVR---RQPDHPAI-----AVNVDTSPQLSPSAHATSTTAAALLM 107
QY 115 KALIGDRGCECPNPVPFSDGLELINFEGACITVEGEINKLVANVAFGRQMLGIHYRF 174
Db 108 GRATGL-----PLPVV-----LVPPMALSRIILGVHYPS 136

QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 10
OS3584 PRELIMINARY; PRT; 416 AA.
AC OS3584;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE BLr7005 protein.
GN BLR7005.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
```

```
QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 9
Q7TW5 PRELIMINARY; PRT; 165 AA.
AC Q7TW5;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Possible conserved transmembrane protein.
GN MB3837C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pcyor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248347; CAD96023.1; -.
KW Complete proteome.
SQ SEQUENCE 165 AA; 17218 MW; 0E084FA26508A4A1 CRC64;

Query Match 10.4%; Score 108; DB 16; Length 165;
Best Local Similarity 24.0%; Pred. No. 0.025;
Matches 46; Conservative 22; Mismatches 52; Indels 72; Gaps 8;

QY 4 GTSHY-----FRLLGAAELAQASCYQKQVHRFARPEALGGTLHNTIAGDLDADF 54
Db 20 GLSHFGHCIGWLILALLGIALPRR---REMLV-----AGAGAFVAHAIA----- 63
QY 55 DISLENDLLKRVAEINAAQNPNEVYLLPQAIQVGSPTSPSGHATONGAPATVL 114
Db 64 -----VLIKLVR---RQPDHPAI-----AVNVDTSPQLSPSAHATSTTAAALLM 107
QY 115 KALIGDRGCECPNPVPFSDGLELINFEGACITVEGEINKLVANVAFGRQMLGIHYRF 174
Db 108 GRATGL-----PLPVV-----LVPPMALSRIILGVHYPS 136

QY 175 DGIQGLLGETI 186
Db 137 DVAVGVALGATV 148

RESULT 10
OS3584 PRELIMINARY; PRT; 416 AA.
AC OS3584;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE BLr7005 protein.
GN BLR7005.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
```

```
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RL Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR ENBL; AP005960; BAC52270.1; -.
GO; GO:0004601; F:peroxidase activity; IEA.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR003320; Cl_perox.
DR Pfam; PF02328; Cl_perox; 1.
KW Complete proteome.
SQ SEQUENCE 416 AA; 4988 MW; F01F13A3E391051C CRC64;

Query Match 10.0%; Score 104; DB 16; Length 416;
Best Local Similarity 27.1%; Pred. No. 0.21;
Matches 42; Conservative 23; Mismatches 70; Indels 20; Gaps 6;

QY 45 TIAGD--LDADFDISLLENDELKRVAEINAAQNPNNEVYLLPQAIQVG-SPTHISYPS 101
Db 267 SIAGNDALVAVFQAKYHN--LWRPITAIRNADLTSPNATPRDFSWLPLGETPMHPEYPC 324
QY 102 GHATQNGAFATVVKALIGLDRGCECFNPFVPSDDGLELINFEGACITVE-GBINKLAVN 160
Db 325 AHCTSAATSTVLQSVVG-DFGESLTSSAPG-----VTRKWSRLQYSDE 370
QY 161 VAFGRQMLGHYRFDGIQGLLGETITVRTLHQEL 195
Db 371 VSNARIWAGHYRFEVGVKMGKIGALTVAQQL 405

RESULT 11
Q7UVW2 PRELIMINARY; PRT; 706 AA.
AC Q7UVW2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Vanadium chloroperoxidase (EC 1.11.1.10).
GN CPO OR RB2401.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleutner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR ENBL; BX294137; CAD72609.1; -.
KW Oxidoreductase; Peroxidase; Complete proteome.
SQ SEQUENCE 706 AA; 76527 MW; 9512A7A54206A6F8 CRC64;

Query Match 9.5%; Score 99; DB 16; Length 706;
Best Local Similarity 28.1%; Pred. No. 1.3;
Matches 45; Conservative 18; Mismatches 67; Indels 30; Gaps 8;

QY 39 GGTLHNTIAGDLADFD--ISLLENDELKRVAEINAAQNPN--NEVYLLPQAIQVGSPT 95
Db 472 GEELRWPVTAIQSGDFDGNLTGSD-----ADWTALGAPDGGDIDVGFTFQ----- 517
QY 96 HPSYPGSHATQNGAFATVVKALIGLD-----GGE-----CFNPNVPSDDGLELINFEGA 146
Db 518 FPTYISGHATFGGALFGLTQEFYGTDDISFTVASEEILELLDNPQLQEAQGL---NLDDA 574
QY 147 CLVYEGEINKLVNVAFGQMLGHYRFDGIQGLLGETI 186
Db 575 ERTFSSFSSEAWEN-GRSRYVLGLIHFDLVDVGQVQSI 613
```

RESULT 12

```
Q825G8 PRELIMINARY; PRT; 525 AA.
AC Q825G8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SAV7489.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
DR ENBL; AP005050; BAC75200.1; -.
GO; GO:0004143; F:diacylglycerol kinase activity; IEA.
GO; GO:0007205; P:protein kinase C activation; IEA.
DR InterPro; IPR008934; AcPase_VanPerase.
DR InterPro; IPR001206; DAGK.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF00781; DAGK; 1.
DR Pfam; PF01569; PAP2; 1.
DR SMART; SM00014; acidPPc; 1.
DR SMART; SM00046; DAGK; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 525 AA; 55271 MW; 38C21722F7996F17 CRC64;

Query Match 9.0%; Score 93; DB 16; Length 525;
Best Local Similarity 27.7%; Pred. No. 3.3;
Matches 28; Conservative 12; Mismatches 31; Indels 30; Gaps 3;

QY 84 LLPQAIQVGSPTH-PSYPGSHATQNGAFATVVKALIGLDRGCECFNPNVPSDDGLELIN 142
Db 90 LVPIRHLTRQPTHTTSPGSHSASAAAFATG-----VALESTG----- 127
QY 143 FEGACITVEGINKLVNVAFGQMLGHYRFDGIQGLLIG 183
Db 128 -----YCALVAPLAAAVAFSRVYVGVHVPDVLGMAIG 161

RESULT 13
Q94AY8 PRELIMINARY; PRT; 299 AA.
AC Q94AY8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AC1G05140/YUP8H12.25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
```



```
Qy 97 -PSYPSGHA-TONGAFATVLKALIGLDRGGECFENPVPPSD--DGLBLINFE----- 144
Db 62 ISAEFQTHQDLNSQSFRRVAKS-----DRSKDLIKSAILDNDFMKNLEWSQIQEIVDCWYP 117
Qy 145 -----GACLTVEGEINKLAVNVAEGROMLGIHYRFDGIQ-----GLLLGE----- 184
Db 118 VDYDKNSCIIKEGDDVGSLLVYVMEDEK----VEVTKEGLKLCCTMGPGKVFGEALILYNCTR 173
Qy 185 TITVRTL 191
Db 174 TATVRTV 180
```

Search completed: September 19, 2004, 03:58:50
Job time : 39.7002 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 19, 2004, 03:59:53 ; Search time 1594.65 Seconds
(without alignments)
3707.849 Million cell updates/sec

Title: US-10-691-383-2_COPY_435_632
Perfect score: 1039
Sequence: 1 VNFGTSHYFLRIGAAELAQ.....GLLGLTITVRLHQLMTF 198

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10691383/runat.17092004.102656.1753/app.query.fasta_1.1230
-DB=EST -QPMT=fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR=score=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10691383 @CGN 1 1 4591 @runat.17092004.102656.1753 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	177.5	17.1	503	9	AU187248	AU187248 AU187248
2	177.5	17.1	538	9	AU191942	AU191942 AU191942
3	113	10.9	369	14	CD063088	CD063088 MA1-00290
4	94.5	9.1	794	12	BI761471	BI761471 603046604
5	93.5	9.0	558	10	AW661332	AW661332 833004D10
6	93	9.0	1207	28	BZ579009	BZ579009 mah2 6117
7	92.5	8.9	745	14	CB037634	CB037634 TC ad2 36
8	92	8.9	818	14	CB977242	CB977242 CAB400003
9	91.5	8.8	537	14	CF192410	CF192410 14f06j2.E
C 10	90.5	8.7	615	14	CA239663	CA239663 SCBFPL506
11	88.5	8.5	604	14	CB925226	CB925226 ABA1 20 E
12	88.5	8.5	612	13	BQ810389	BQ810389 1030017B0
13	88.5	8.5	629	14	CB331374	CB331374 3529 1 36
14	88.5	8.5	639	14	CD223063	CD223063 CCC1_25 C
15	88.5	8.5	666	14	CD229795	CD229795 CCC1_19 D
16	88.5	8.5	755	12	BI772791	BI772791 603054341
C 17	87.5	8.4	743	14	CF571423	CF571423 MCS013F01
18	87.5	8.4	780	14	CK025517	CK025517 AGENCOURT
C 19	87.5	8.4	944	28	CC327659	CC327659 OSUEP65TH
20	87	8.4	599	14	CF485754	CF485754 POL1_33 A
21	87	8.4	703	12	BJ272534	BJ272534 BJ272534
22	86.5	8.3	468	9	AL045228	AL045228 DKF2p434H
23	86.5	8.3	577	13	BQ829915	BQ829915 L6in2239
C 24	86.5	8.3	688	12	BJ598040	BJ598040 BJ598040
C 25	86.5	8.3	718	12	BJ600919	BJ600919 BJ600919
26	86	8.3	710	13	BW208048	BW208048 BW208048
C 27	86	8.3	3218	11	AK088408	AK088408 Mus muscu
C 28	86	8.3	3261	29	AY408110	AY408110 Mus muscu
C 29	86	8.3	4324	11	AK085707	AK085707 Mus muscu
C 30	85.5	8.2	568	13	BQ831943	BQ831943 L6in1122
31	85.5	8.2	592	10	BE382065	BE382065 601272225
C 32	85.5	8.2	620	28	BZ332752	BZ332752 hx30008.b
33	85.5	8.2	629	10	BE374596	BE374596 601225384
34	85.5	8.2	726	12	BI648082	BI648082 603279290
C 35	85.5	8.2	744	12	BJ598045	BJ598045 BJ598045
36	85.5	8.2	773	12	BI157032	BI157032 602921306
37	85.5	8.2	785	12	BI688650	BI688650 603311262
38	85.5	8.2	787	12	BI156730	BI156730 602922150
39	85.5	8.2	844	12	BI158626	BI158626 602922395
40	85.5	8.2	887	14	CA980253	CA980253 AGENCOURT
C 41	85.5	8.2	903	10	BF311105	BF311105 601898555
42	85.5	8.2	1065	12	BM924437	BM924437 AGENCOURT
43	85.5	8.2	1988	11	BC023725	BC023725 Mus muscu
44	85.5	8.2	3110	11	AK031542	AK031542 Mus muscu
45	85.5	8.2	3372	11	BC046553	BC046553 Mus muscu

ALIGNMENTS

RESULT 1	AU187248	503 bp	mRNA	linear	EST 14-OCT-2003
LOCUS	AU187248	Porphyra yezoensis TU-1	sporophytes	Porphyra yezoensis	
DEFINITION	CDNA clone PF010b07_r 5'	mRNA sequence.			
ACCESSION	AU187248				
VERSION	AU187248.1	GI:31920687			
KEYWORDS	EST.				
SOURCE	Porphyra yezoensis				
ORGANISM	Porphyra yezoensis				
	Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;				
REFERENCE	1 (bases 1 to 503)				


```

Qy 82 ThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSer 101
Db 525 CCTCAGCTGCTGGAGGAGGGGGTGGGGCGGGGGCTCCG----- 563
Qy 102 GlyHisAlaThrGlnAenGlyAlaPheAlaThrValLeuLysAlaLeuIleGlyLeuAsp 121
Db 564 -----GGCAGTGCTGGCGGGCTCTGGCTGCCCTGGCTGGACCATGTC 605
Qy 122 ArgGlyGlyGluCysPhe-----ProAsnProValPhe 132
Db 606 AGGAGCGGGTCTTGCTCCACGCGCTTGGCGAGCGCTCAGTAC 647

RESULT 5
LOCUS BZ579009 1207 bp mRNA linear EST 01-MAY-2000
DEFINITION msh2_6117.xl C. reinhardtii CC-125 -S, Lambda Zap II Chlamydomonas
reinhartdii cDNA, mRNA sequence.
ACCESSION BZ579009.1 GI:27214070
VERSION BZ579009.1
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 558)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants
JOURNAL Unpublished (2000)
COMMENT Contact: Elizabeth H. Harris
DUMB Box 91000
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.

FEATURES
source
1. 558
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-125 wild type mt+ 137c"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-125 -S, Lambda Zap II"
/notes="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Strain CC-125 was grown to mid-log phase in TAP
medium, washed twice in TAP-S (sulfur-deficient TAP
medium) and incubated in TAP-S. RNA was isolated after 1,
2 and 4 hours. An equal amount of RNA from each time point
was pooled and polyA mRNA purified. cDNA was synthesized
and directionally cloned into lambda ZAP II (Stratagene)
in the EcoRI (5') and XhoI (3') sites. pBluescript II SK-
plasmids were excised from the lambda ZAP clones by
superinfection with ExAssist (Stratagene) phage."

ORIGIN
Alignment Scores:
Pred. No.: 2.01 Length: 558
Score: 93.50 Matches: 39
Percent Similarity: 30.53% Conservative: 19
Best Local Similarity: 20.53% Mismatches: 61
Query Match: 9.00% Indels: 71
DB: 10 Gaps: 6

US-10-691-383-2_COPY_435_632 (1-198) x AW661332 (1-558)

Qy 6 SerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCysTyrGln 25
Db 34 GCCCGGTCTTCAAGCTGCGGCTCTTCGTTCTTTGACCGCCGCGGCTGCTTCGT 93
Qy 26 LysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyThrLeuHisAsnThr 45

```

```

Db 94 ATCAAGTACGGGAGCTCTTTCGGCGCCCATCACCGCATCCGACCCATGGC--- 150
Qy 46 IleAlaGlyAspLeuAspAlaPheAspIleSerLeuLeuGluAsnAspGluLeuLeu 65
Db 151 ---GTCGGCAGCCGCTGACCCACCTGG----- 174
Qy 66 LysArgValAlaGluLeuAsnAlaAlaGlnAenProAsnAsnGluValThrTyrLeuLeu 85
Db 175 -----ACGCGCTGCTG 186
Qy 86 ProGlnAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThr 105
Db 187 -----GCCACCCCGCCGAGTTCCTCCCTCGGCGCACCATGTC 228
Qy 106 GlnAsnGlyAlaPheAlaThrValLeuLysAlaLeuIleGly----- 119
Db 229 AGTCGGGCGCTGGACTGCATGGGGGGGGCATCTTCGGCGCCCTGGGCGCCCTGAAC 288
Qy 120 -----LeuAspArgGlyGlyGluCysPheProAsnProValPhePro 133
Db 289 GTAACCACTACGGCGCCCTGGACATCTCGCCCGCACCTTCCCT----- 333
Qy 134 SerAspGlyLeuLeuLeuLeuAsnPheGluGlyAlaCysLeuThrTyrGluGlyGlu 153
Db 334 -----AGCTTCGGCTCGCT----- 348
Qy 154 IleAsnLysLeuAlaValAenValAlaPheGlyArgGlnMetLeuGlyIleHisTyrArg 173
Db 349 -----GCCGTGAGTGGCGTACAGCGCTGTACGCTGTGTCACCTTCAAC 396
Qy 174 PheAspGlyIleGlnGlyLeuLeuLeuGly 183
Db 397 AGTCCAACTGACGAGTCAATCTCGGC 426

RESULT 6
LOCUS BZ579009
DEFINITION msh2_6117.xl msh Pseudomonas aeruginosa genomic clone msh2_6117,
genomic survey sequence.
ACCESSION BZ579009.1 GI:27214070
VERSION BZ579009.1
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 1207)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) in press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1. 1207
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="M5H"
/db_xref="taxon:287"
/clone="msh2_6117"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."

ORIGIN

```



```

Pred. No.: 3.24 Length: 537
Score: 91.50 Matches: 39
Percent Similarity: 35.26% Conservative: 22
Best Local Similarity: 22.54% Mismatches: 50
Query Match: 8.81% Indels: 62
DB: 14 Gaps: 6

US-10-691-383-2_COPY_435_632 (1-198) x CF192410 (1-537)

QY 3 PheGlyThrSerHisTyrPheArgLeuIleGlyAlaAlaGluLeuAlaGlnAlaSer 22
DB 47 TTCTCTAGCACCGGTTCTTCTGTTACAGTTTCGGCCACCACTGCCACTCATGGACTAGT 106
QY 23 CyTyTyrGlnLysTrp---GlnValHisArgPhe----- 32
DB 107 CAGAACCAACGCTGGTAACAGCCCATAGATACCTCGCATCGGGCCGCCACCTTCAG 166
QY 33 -----AlaArg-ProGluAla----- 37
DB 167 GTGTTTACGCCACCAAGCAGGCAAGGTCCAAGACCTCTCATCGCTCTTGTGAAGCA 226
QY 38 -----LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAs 49
DB 227 ACAAGAAGGTGAGTGAACCGATCTTGGGGCAAGCTCATGATACCGTAGCTGATGA 286
QY 49 pLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAl 69
DB 287 TCTCAGCGTTAACAACGTAGTGAAGGTTAATGCGCTGAGTGAGGATCCAGATCAGCGCTTGGC 346
QY 69 aGluIleAsnAlaAlaGlnAsnProAsnGluAluValThrTyLeuLeuProGlnAlaI 89
DB 347 CGATGTAATCAAGATGATCATGGAGGAGAGAGGTAGAACAGCCTTGAGTAGGGCGAG 406
QY 89 eGlnValGlySerProThrHisProSerTyProSerGlyHisAlaThrGlnAsnGlyAl 109
DB 407 CTCTGTG-----AACGGAGG 421
QY 109 aPheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGly----- 123
DB 422 CTTTTCGACC-----TTTGGCGAAGACACGGTCCAGACGCCCAAGATG 466
QY 124 -----GlyGluCysPheProAsnProValPhePro 133
DB 467 CAAGATCGGCAACTGCGAACCAAGCCCACTTGTGTC 503

RESULT 10
CA239663/c 615 bp mRNA linear EST 25-SEP-2003
LOCUS SCBFPL5067C03.b Saccharum officinarum FL5 Saccharum officinarum
DEFINITION cDNA clone SCBFPL5067C03 3', mRNA sequence.
ACCESSION CA239663
VERSION CA239663.1 GI:35312481
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 615)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenhariaia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 067 row: C column: 03

```

Seq primer: SP6 Promoter primer.
Location/Qualifiers
1. .615
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCBFPL5067C03"
/lab_host="DH10B"
/notes="Organ: Developed inflorescence (20cm-long) without rachis; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Developed inflorescence (20cm-long) without rachis]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.lad.ic.unicamp.br/public>"

Alignment Scores:
Pred. No.: 5.31 Length: 615
Score: 90.50 Matches: 31
Percent Similarity: 44.25% Conservative: 19
Best Local Similarity: 27.43% Mismatches: 36
Query Match: 8.71% Indels: 27
DB: 14 Gaps: 4

ORIGIN
US-10-691-383-2_COPY_435_632 (1-198) x CA239663 (1-615)
QY 36 GluAlaLeuGlyGlyThrLeuHisAsnThrIle---AlaGlyAspLeuAspAlaAspPhe 54
DB 571 GATGGCATTTGGGGAACTTCCAGCACTATAGGATCTGGGGTCTGGATGCACCTCCA 512
QY 55 AspIleSerLeuLeuGluAsnAspGluLeuLeuLysArgValAlaGluIleAsnAlaAla 74
DB 511 GAAATTTCCCTTCCCAACCTGAGGTG-----GTGGTGGCAACACAGCAATC 464
QY 75 GlnAsnProAsnAsnGluValThrTyLeuLeuProGlnAlaIleGlnValGlySerPro 94
DB 463 ACACAGCCTTAACAAAGACATGGAGTATGTTGCTCCACCAAT----- 419
QY 95 ThrHisProSerTyProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeu 114
DB 418 -----GAAGATGCTGCACCTACAGCAACCTT 392
QY 115 LysAlaLeuIleGlyLeu-----AspArgGlyGlyGluCysPhePro 128
DB 391 GAACATGCTGTGGCATACTGGAAGTTACACCTGCTGCTGGAGGATGTTGCACCT 332
QY 129 AsnProValPheProSerAspAspGlyLeuGluLeuIle 141
DB 331 TCTGCTCGGAGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

RESULT 11
CB925226 604 bp mRNA linear EST 28-APR-2003
LOCUS ABAL_20_E07.g1_A012 Abscisic acid-treated seedlings Sorghum bicolor
DEFINITION cDNA clone ABAL_20_E07_A012 5', mRNA sequence.
ACCESSION CB925226
VERSION CB925226.1 GI:30161497
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 604)
Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Buchanan,C.D., Eastman,A. and Pratt,L.H.

TITLE	COMMENT
An EST database from Sorghum: ABAL-treated seedlings Unpublished (2003)	
Other_ESTs: ABAL_20_E07.bl_A012	
Contact: Cordonnier-Pratt MM	
Laboratory for Genomics and Bioinformatics	
The University of Georgia, Department of Plant Biology	
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA	
Tel: 706 542 1860	
Fax: 706 583 0210	
Email: mmpratt@uga.edu	
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.	
seg primer: SUGS (CTTCTGCTCTAAAAAGTCGCG).	

```

FEATURES             Location/Qualifiers
     source            1..604
                     /organism="Sorghum bicolor"
                     /mol_type="mRNA"
                     /cultivar="IS3620C"
                     /db_xref="taxon:4558"
                     /clone="ABAI_20 E07 A012"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="Abscisic acid-treated seedlings"
                     /notes="Vector: pME18S-FL3; Site_1: XhoI; Site_2: XhoI; The
library was prepared from polyA+ RNA from seedlings grown
in hydroponic culture. After 12 days, medium was
supplemented with 1 mM abscisic acid (ABA), while leaves
were misted with a solution of 1 mM ABA. Roots and leaves
were harvested after 3, 6, 12, and 24 hr and material from
all time points was combined prior to RNA isolation.
Double-stranded cDNA was cloned unidirectionally into
different DraIII sites of the pME18S-FL3 vector (5-prime
DraIII site is CACTGTTGTG, 3-prime DraIII site is
CACCAATGTG). XhoI excises the cDNA insert."
ORIGIN

```

```

ORIGIN
Alignment Scores:
Pred. No.:      8.86      Length:      604
Score:          88.50     Matches:     47
Percent Similarity: 39.31%  Conservative: 21
Best Local Similarity: 27.17% Mismatches:   51
Query Match:      8.52%   Indels:      54
DB:               14      Gaps:       11

US-10-691-383-2_COPY_435_632 (1-198) x CB925226 (1-604)

QY      11  LeuIleGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29
Db      88  CTTTCGGCGTGCCTCTGCGGGGAGCGGGGGCTGTTCCTTCACCCCTCCCGCGTG 147
QY      30  HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly 48
Db      148 GTCCAACTCTCTGGCGGCCCGCCAGGGCGTTC-----CTGCACAAGGGTTTCCTGTGC 198
QY      49  AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLys---Arg 67
Db      199 GATCGCGAGTGCAC---CACCTCATCGTGTGGCCCAAGGACAAACTGGAGAAAGTCTATG 255
QY      68  ValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnLuvAl----- 81
Db      256 GTACGGGACAAACGAGTCCGCTTAAGAGCGTCCAGAGCGAGGTGGCGCACCGACCTCGGCATG 315
QY      81  ----- 81
Db      316 TTCTGTGGAGAAGACAGGATGATGACAGTAAAGGGGAATAGAGGAGGATAGCTGTTGG 375
QY      82  ThrTyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySer 93
Db      93  ----- 93

```

376	Db	ACATTTCTTCCACGAGAAATGGTGAATCCATTCCAGATATTACCTACCAAGATGGTGAG	435
94	Qy	ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla	109
436	Db	AAATACGAACCCCATCTACTACTTCCATGTACAAAATAATCAAGCCCTGNGTGCCAT	495
110	Qy	---PheAlaThrValLeuIysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhe	127
496	Db	CGCATTCGCCACTGTCTCATGTACTTATCTAATGTTGAGAAGGTGGAGACCATCTTTC	555
128	Qy	ProAsnPro-----ValPheProSerAspAsp	136
556	Db	CCCAATGCAGAGGGGAAGCTATTACACCCAGGACGAC	594
<p> RESULT 12 BQ810389 LOCUS DEFINITION 1030017304.y2 C. reinhardtii CC-1690, Deflagellation (normalized), lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence. BQ810389 ACCESSION VERSION KEYWORDS SOURCE ORGANISM Chlamydomonas reinhardtii Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas. 1 (bases 1 to 612) Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030 Unpublished (2002) Contact: Charles Hauser DCMB Box 91000 Duke University Durham, NC 27708-1000 Tel: 919 613 8159 Fax: 919 613 8177 Email: chauser@duke.edu. Location/Qualifiers 1..612 /organism="Chlamydomonas reinhardtii" /mol_type="mRNA" /strain="CC-1690 wild type mt+ 21gr" /db_xref="taxon:3055" /clone_lib="C. reinhardtii CC-1690, Deflagellation (normalized), lambda zap II" /notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Deflagellation library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells which had been re-synthesizing flagella for 15, 30 and 60 min after being deflagellated by pH shock. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II. (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806." </p>			
<p> ORIGIN Alignment Scores: Pred. No.: 9.06 Length: 612 Score: 88.50 Matches: 45 Percent Similarity: 73.36% Conservative: 23 Best Local Similarity: 24.73% Mismatches: 73 Query Match: 8.52% Indels: 41 DB: 13 Gaps: 11 </p>			
<p> US-10-691-383-2_COPY_435_632 (1-198) x BQ810389 (1-612) Qy 4 GlyThrSerHisTyrPhe---ArgLeuIleGlyAlaGluLeuAlaGlnAraGalaSer 22 </p>			

```

Db      145 GCGGTGAAGCCCTCTTCGCGAGAGCGGCTGGCGCTGGCGGACCGCGTGGAGAGGAGCAAG 204
Qy      23 CysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeuGlyGlyThrLeu 42
Db      205 GCCTACATGGTGTACCGGTGGAGGAGCGCGAGCGCGAG---CTCAACGGCTTCTC 261
Qy      43 -----HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSer 57
Db      262 AAGGAGCTGGAGGCCCATCGGCACACGCTGGGCGATCACAGACGTGCGAGCTGAGCGTACC 321
Qy      58 LeuLeuGluAsnAspGluLeu---LeuLysArgValAlaGluIleAsn---AlaAlaGln 75
Db      322 AGCTTGAGGAGGTTCCTCAACATCGCGCGCGCGCGGAGGTGGAGCGCGCGCGCAGC 381
Qy      76 AsnProAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 95
Db      382 AGCGGCACACCGAGGTCAACGACGCTGTGGACGCGCGCGCTCAAGATCCCGGTG 441
Qy      96 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 115
Db      442 GCGCGCGAGGTG---GTGGGACACCCACCACCGCGCGCGCTAC-----CGC 486
Qy      116 AlaLeuIleGlyLeuAspArgGlyGly-----GluCys-----PhePro 128
Db      487 GTGGATGGGACCGGACGAGCGCGCGCTGATCATCATGAGTGGAGCTGCGAGCTGCCG 546
Qy      129 AsnProValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeu 148
Db      547 ---CCAGCTCCCGCGAGCGCGAGCTGAG-----576
Qy      149 ThrTyrGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeu 168
Db      577 -----GTCGCGCGCGCGCGCGAGGCGAGGTGATG 606
Qy      169 GlyIle 170
Db      607 GCGGTG 612

RESULT 13
LOCUS   CB331374
DEFINITION 3529_136_1_G07.y_1 3529 - 2 mm ear tissue from Schmidt and Hake
labs Zea mays cDNA, mRNA sequence.
ACCESSION CB331374
VERSION   CB331374.1 GI:28910308
KEYWORDS EST.
SOURCE   Zea mays
ORGANISM Zea mays
REFERENCE 1 (bases 1 to 629)
AUTHORS Walbot,V.
TITLE    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 629)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3529_136_1 row: G column: 07.
Location/Qualifiers
1. 629
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="ear"
/dev_stage="2 mm"
FEATURES
source
location(136_1)

```

```

/lab_host="E. coli XL0LR"
/clone_lib="3529 - 2 mm ear tissue from Schmidt and Hake
labs"
/note="Organ: ear; Vector: PAD-GAL4-2.1; Site 1: EcoRI;
Site 2: XhoI; RNA isolated by Hake lab. 1 million pfu
amplified. Ampicillin is the selection marker."

ORIGIN
Alignment Scores:
Pred. No.: 9.48 Length: 629
Score: 88.50 Matches: 46
Percent Similarity: 38.37% Conservative: 20
Best Local Similarity: 26.74% Mismatches: 57
Query Match: 8.52% Indels: 49
DB: 14 Gaps: 9

US-10-691-383-2_COPY_435_632 (1-198) x CB331374 (1-629)

Qy      11 LeuIleGlyAlaAlaGluLeuAlaGlnArg---AlaSerCysTyrGlnLysTrpGlnVal 29
Db      79 CTCACCGGATCGCGCTTCTGGCGGAGCGCGGCGCTGTTCTTCGACCCCATCCCGGTG 138
Qy      30 HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly 48
Db      139 GTTCAACTCTCTTGGCGCGCGCGCGCTTC-----TTGCACAGGGTTTCTGTGTG 189
Qy      49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLysArgVal 68
Db      190 GATCGGAGTGGCGACCACTCATCGCGCTGGCCAAAGGCAAGCTGGAGAGTCTATGTGTG 249
Qy      69 AlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal-----81
Db      250 GCGGACAAACAAGTCCGCTAAGAGTGTCCAGAGTGGTGGCGCACAGTTCGCGCATGTTC 309
Qy      82 -----Thr 82
Db      310 CTCGAGAAGAAGCAGGATGAAGTAGTAACGAGGATAGAGGAGGATATCTGCTTGGACA 369
Qy      83 TyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySerPro 94
Db      370 TTCTTCCACAGAGAATGTTGAAGCATTCAGATATTACACTACCAGAAATGGTGAGAAA 429
Qy      95 ThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla-----109
Db      430 TATGAACCCCACTACGACTACTTCCATGACAAAATAATCAGGCCCTGGGCGGCATCGC 489
Qy      110 PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhePro 128
Db      490 ATTGCCACTGTGTGATGTACCTATCAATGTGAGAAGGGTGGAGAGACCATCTTCCCC 549
Qy      129 AsnPro-----ValPheProSerAspAsp 136
Db      550 AATGCAGAGGGGAAGCTATTGCAACCAAGGACGAC 585

RESULT 14
LOCUS   CD223063
DEFINITION CCC1_25_C01_91_A007 Callus culture/cell suspension Sorghum bicolor
cDNA clone CCC1_25_C01_A007 5', mRNA sequence.
ACCESSION CD223063
VERSION   CD223063.1 GI:30966497
KEYWORDS EST.
SOURCE   Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE 1 (bases 1 to 639)
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 639)
Cordonnier-Pratt,M.-M., Wentzel,V., Suzuki,Y., Sugano,S.,
Klein,R.R., Liang,C., Sun,F., Sullivan,R., Shah,M., Rathore,K.,
Eastman,A. and Pratt,L.H.
An EST database from Sorghum: callus culture and cell suspension
Unpublished (2003)
JOURNAL

```



```
Db 100 CTGCGCGCTGCGCTCTGCGGCGAGCGCGGGCGCTGGTTCCTTCGACCCCTCCCGCGTG 159
Qy 30 HisArgPheAla---ArgProGluAlaLeuGlyGlyThrLeuHisAsnThrIleAlaGly 48
Db 160 GTCCAACTCTCGTGGCGCCCGCGGCGCTTC-----CTGCACAAGGGTTTCTCTGTCG 210
Qy 49 AspLeuAspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLys---Arg 67
Db 211 GATCGCGAGTGCAC---CACCTCATCTGCTGGCCCAAGGACAACTGGAGAAAGTCTATG 267
Qy 68 ValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluVal----- 81
Db 268 GTAGCGGACAAACGAGTCCGGTAGAGCGTCCAGAGCGAGGTGCGCACCGATCTCCGGCATG 327
Qy 81 ----- 81
Db 328 TTCTGTGAGAGGAAGCAGGATGAAGTAGTAAGGGGAATAGAGGAGAGGATAGCTGCTTGG 387
Qy 82 ThrTyrLeuLeuPro-----GlnAlaIleGlnVal-----GlySer 93
Db 388 ACATTTCTCCACGAGAGAATGGTGAATCCATTCCAGATATTACACTACCAGAAATGGTGAG 447
Qy 94 ProThrHisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAla----- 109
Db 448 AAATACGAACCCCACTATGACTCTCCATGACAAAAAATAATCAAGCCCTGGGGTGGCCAT 507
Qy 110 ---PheAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlu---CysPhe 127
Db 508 CGCATTCGCACTGTGCTCATGTACTTATCTAATGTTGAGAGGGTGGAGAGACCATCTTC 567
Qy 128 ProAsnPro-----ValPheProSerAspAsp 136
Db 568 CCCAATGCAGAGGGGAAGCTATTACAGCCCAAGGACGAC 606
```

Search completed: September 19, 2004, 09:22:46
Job time : 1601.65 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 19, 2004, 01:14:52 ; Search time 23.9771 Seconds
(without alignments)
1468.042 Million cell updates/sec

Title: US-10-691-383-2

Perfect score: 3528

Sequence: 1 MLCHAADTRGSPMPDTGVL.....SIDGMCNGLVYTGVDCA 676

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	2514	71.3	557	1	PRXV_ASCNO
2	114.5	3.2	690	1	NRF_HUMAN
3	113.5	3.2	912	1	CAD2_CHICK
4	110.5	3.1	439	1	COMP_BOVIN
5	107	3.0	1819	1	GCP6_HUMAN
6	106.5	3.0	947	1	TOPI_MYCLE
7	106.5	3.0	1016	1	DPOL_RHILE
8	106	3.0	303	1	LPXC_PSESM
9	106	3.0	426	1	HISX_METH
10	106	3.0	3670	1	CSM3_HUMAN
11	104.5	3.0	554	1	SECD_RHOCA
12	104.5	3.0	615	1	MUTA_MYCTU
13	104.5	3.0	1305	1	FTSK_YERPE
14	103.5	2.9	755	1	COMP_MOUSE
15	103.5	2.9	755	1	COMP_RAT
16	103.5	2.9	757	1	COMP_HUMAN
17	102.5	2.9	469	1	GLNA_VIBCH
18	102	2.9	737	1	PL02_HUMAN
19	101.5	2.9	567	1	SYFB_HALN1
20	101.5	2.9	644	1	FTSH_ECO57
21	101.5	2.9	644	1	FTSH_ECOLI
22	101.5	2.9	906	1	CAD2_MOUSE
23	101.5	2.9	1036	1	HP12_DEIRA
24	101	2.9	741	1	HLVA_VIBCH
25	101	2.9	877	1	CAD2_BOVIN
26	101	2.9	1744	1	TENS_CHICK
27	100	2.8	566	1	ATFI_SCHPO
28	100	2.8	625	1	DXS_HAEIN
29	100	2.8	877	1	SECA_GUITH
30	100	2.8	948	1	HP11_DEIRA
31	100	2.8	2124	1	PGCA_RAT
32	100	2.8	2319	1	NTC3_RAT
33	99.5	2.8	791	1	TEX_BORPE

RESULT 1

ID	PRXV_ASCNO	STANDARD;	PRT;	557 AA.
AC	P81701;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Vanadium haloperoxidase (EC 1.11.1.-) (V-BPO).			
OS	Ascopyllium nodosum (knotted wrack) (Brown seaweed).			
OC	Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaea;			
OC	Ascopyllium.			
OX	NCBI_TaxID=52969;			
RN	[1]			
RP	SEQUENCE, X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS), AND FUNCTION.			
RX	MEDLINE=20013071; PubMed=10543953;			
RA	Weyand M., Hecht H.-J., Kiess M., Liaud M.-F., Vilter H.,			
RA	Schomburg D.;			
RT	"X-ray structure determination of a vanadium-dependent			
RT	haloperoxidase from Ascopyllium nodosum at 2.0-A resolution."			
RL	J. Mol. Biol. 293:595-611(1999).			
RN	[2]			
RP	SEQUENCE OF 320-556 FROM N.A., SEQUENCE OF 326-341; 383-426; 471-479			
RP	AND 481-556, AND FUNCTION.			
RX	MEDLINE=96081028; PubMed=8564812;			
RA	Vilter H.;			
RT	"Vanadium-dependent haloperoxidases."			
RL	(in) Sigel H., Sigel A. (eds.);			
RL	Metal ions in biological system-vanadium and its role in life,			
RL	pp.31:325-362, Marcel Dekker, New York (1995).			
CC	-1- CATALYTIC ACTIVITY: Halide + H(2)O(2) + H(+) = HOHal + H(2)O.			
CC	-1- COFACTOR: Vanadium.			
CC	-1- SUBUNIT: Homodimer; disulfide-linked.			
CC	-1- SIMILARITY: TO OTHER BACTERIAL NON-HEME BROMO- AND CHLORO-			
CC	PEROXIDASES.			
DR	PDB; 1Q19; 10-JUN-00.			
DR	InterPro; IPR008934; AcPase_VanParase.			
DR	InterPro; IPR000326; PA_PTPase.			
KW	Oxidoreductase; Peroxidase; Vanadium; 3D-structure;			
KW	Pyridinolone carboxylic acid.			
FT	MOD RES 1 1			
FT	DISULFID 3 3			
FT	DISULFID 41 41			
FT	DISULFID 77 86			
FT	DISULFID 441 462			
FT	DISULFID 544 555			
FT	ACT SITE 411 411			
FT	ACT SITE 418 418			
FT	METAL 486 486			
FT	CONFLICT 321 321			
FT	CONFLICT 341 341			
FT	CONFLICT 403 404			
FT	CONFLICT 407 408			
FT	CONFLICT 409 409			
FT	CONFLICT 441 444			
FT	CONFLICT 470 470			
FT	HELIX 15 37			

P78871 schizosacch
Q53114 mycobacteri
Q92fi8 rhodobacter
Q8xg72 salmoneilla
Q921v3 rattus norv
P50851 homo sapien
P27050 bacillus ci
P35442 homo sapien
P21241 brassica na
Q89af2 buchnera ap
Q50596 mycobacteri
Q03043 drosophila

ALIGNMENTS

50 52
53 53
54 53
55 53
56 53
57 53
58 53
59 53
60 53
61 53
62 53
63 53
64 53
65 53
66 53
67 53
68 53
69 53
70 53
71 53
72 53
73 53
74 53
75 53
76 53
77 53
78 53
79 53
80 53
81 53
82 53
83 53
84 53
85 53
86 53
87 53
88 53
89 53
90 53
91 53
92 53
93 53
94 53
95 53
96 53
97 53
98 53
99 53
100 53
101 53
102 53
103 53
104 53
105 53
106 53
107 53
108 53
109 53
110 53
111 53
112 53
113 53
114 53
115 53
116 53
117 53
118 53
119 53
120 53
121 53
122 53
123 53
124 53
125 53
126 53
127 53
128 53
129 53
130 53
131 53
132 53
133 53
134 53
135 53
136 53
137 53
138 53
139 53
140 53
141 53
142 53
143 53
144 53
145 53
146 53
147 53
148 53
149 53
150 53
151 53
152 53
153 53
154 53
155 53
156 53
157 53
158 53
159 53
160 53
161 53
162 53
163 53
164 53
165 53
166 53
167 53
168 53
169 53
170 53
171 53
172 53
173 53
174 53
175 53
176 53
177 53
178 53
179 53
180 53
181 53
182 53
183 53
184 53
185 53
186 53
187 53
188 53
189 53
190 53
191 53
192 53
193 53
194 53
195 53
196 53
197 53
198 53
199 53
200 53
201 53
202 53
203 53
204 53
205 53
206 53
207 53
208 53
209 53
210 53
211 53
212 53
213 53
214 53
215 53
216 53
217 53
218 53
219 53
220 53
221 53
222 53
223 53
224 53
225 53
226 53
227 53
228 53
229 53
230 53
231 53
232 53
233 53
234 53
235 53
236 53
237 53
238 53
239 53
240 53
241 53
242 53
243 53
244 53
245 53
246 53
247 53
248 53
249 53
250 53
251 53
252 53
253 53
254 53
255 53
256 53
257 53
258 53
259 53
260 53
261 53
262 53
263 53
264 53
265 53
266 53
267 53
268 53
269 53
270 53
271 53
272 53
273 53
274 53
275 53
276 53
277 53
278 53
279 53
280 53
281 53
282 53
283 53
284 53
285 53
286 53
287 53
288 53
289 53
290 53
291 53
292 53
293 53
294 53
295 53
296 53
297 53
298 53
299 53
300 53
301 53
302 53
303 53
304 53
305 53
306 53
307 53
308 53
309 53
310 53
311 53
312 53
313 53
314 53
315 53
316 53
317 53
318 53
319 53
320 53
321 53
322 53
323 53
324 53
325 53
326 53
327 53
328 53
329 53
330 53
331 53
332 53
333 53
334 53
335 53
336 53
337 53
338 53
339 53
340 53
341 53
342 53
343 53
344 53
345 53
346 53
347 53
348 53
349 53
350 53
351 53
352 53
353 53
354 53
355 53
356 53
357 53
358 53
359 53
360 53
361 53
362 53
363 53
364 53
365 53
366 53
367 53
368 53
369 53
370 53
371 53
372 53
373 53
374 53
375 53
376 53
377 53
378 53
379 53
380 53
381 53
382 53
383 53
384 53
385 53
386 53
387 53
388 53
389 53
390 53
391 53
392 53
393 53
394 53
395 53
396 53
397 53
398 53
399 53
400 53
401 53
402 53
403 53
404 53
405 53
406 53
407 53
408 53
409 53
410 53
411 53
412 53
413 53
414 53
415 53
416 53
417 53
418 53
419 53
420 53
421 53
422 53
423 53
424 53
425 53
426 53
427 53
428 53
429 53
430 53
431 53
432 53
433 53
434 53
435 53
436 53
437 53
438 53
439 53
440 53
441 53
442 53
443 53
444 53
445 53
446 53
447 53
448 53
449 53
450 53
451 53
452 53
453 53
454 53
455 53
456 53
457 53
458 53
459 53
460 53
461 53
462 53
463 53
464 53
465 53
466 53
467 53
468 53
469 53
470 53
471 53
472 53
473 53
474 53
475 53
476 53
477 53
478 53
479 53
480 53
481 53
482 53
483 53
484 53
485 53
486 53
487 53
488 53
489 53
490 53
491 53
492 53
493 53
494 53
495 53
496 53
497 53
498 53
499 53
500 53
501 53
502 53
503 53
504 53
505 53
506 53
507 53
508 53
509 53
510 53
511 53
512 53
513 53
514 53
515 53
516 53
517 53
518 53
519 53
520 53
521 53
522 53
523 53
524 53
525 53

FT STRAND 528 531
FT TURN 533 534
FT STRAND 537 539
FT TURN 540 541
FT STRAND 542 543
FT HELIX 544 545
SQ SEQUENCE 557 AA; 60343 MW; E3D8557AB92B16F4 CRC64;

Query Match 71.3%; Score 2514; DB 1; Length 557;
Best Local Similarity 86.5%; Pred. No. 1.8e-177;
Matches 480; Conservative 31; Mismatches 40; Indels 4; Gaps 1;

QY 124 QTCNSDDALDPTAPNRDNVAFASRRDARRRRDGTGTCQITNGTDLATMFKSLPH 183
DB 1 QTCSTDDADDPTPNRRDDEAFASRAAKRELEGTGTCQINNGETDLAAFKSLPH 60

QY 184 DELQVTTADDFAILDCILNGDFSI CEDVPA----GDPAGRLVNPATAFAIDISGPAPSA 239
DB 61 DDLGQVDADAFALDCLNGDLISICEDVPVGNSEGDVPVGRVNPATAFAIDISGPAPSA 120

QY 240 TTTPPVPTLSSPELAQAQLAEVYMWALARDVPFMOYGTDEITTTAAANLAGMGPPNLDV 299
DB 121 TTTPPVPTLPSPELAQAQLAEVYMWALARDVPFMOYGTDDITVTAANLAGMGPPNLDV 180

QY 300 SIGSDGTVDPSQLFRATFVGTVETGPFVSQLLVNSFTIDAITYEPKQETAPDLNVMVDF 359
DB 181 SIGSDGTVDPSQLFRATFVGTVETGPFVSQLLVNSFTIDSITYEPKQETAPDLNVMVDF 240

QY 360 DEWLNTQNGGPPAGPDELDELFI RNARDLARVSFDVNTINTAYRGSLLLELGAFSRP 419
DB 241 DEWLNTQNGGPPAGPDELDELFI RNARDLARVSFDVNTINTAYRGSLLLELGAFSRP 300

QY 420 GINGPFTDSROAGFNVFGTSHYFRILGAEALQASQYKQWQVHRPAREALGGTLHNT 479
DB 301 GVGPFIDIDRQAGFNVFGTSHYFRILGAEALQASQYKQWQVHRPAREALGGTLHNT 360

QY 480 IAGDLADFDISLLENDELKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSPSGHAT 539
DB 361 IKGELNADFDLSLENAELLKRVAEINAAQNPNEVYLLPQAIQVGSPTHPSPSGHAT 420

QY 540 QNGAFATVTKALIGDRGCGCFPNPVPDDGLLEINFEACLTYPGEINKLAVNVAAGR 599
DB 421 QNGAFATVTKALIGDRGCGCFPNPVPDDGLLEINFEACLTYPGEINKLAVNVAAGR 480

QY 600 QMLGIHYRFDGIQGLLIGETITVTRTHOELMTAREATFEFRLTGTVIKLFDGTFSD 659
DB 481 QMLGIHYRFDGIQGLLIGETITVTRTHOELMTAREATFEFRLTGTVIKLFDGTFSD 540

QY 660 GDMCSGLVYTGVDAC 674
DB 541 GFKCPGLVYTGVCNC 555

RESULT 2
NRF_HUMAN
ID NRF_HUMAN STANDARD; PRT; 690 AA.
AC O15226; Q9UJ91;
DT 15-JUL-1998 (Rel. 36, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE NF-kappa-B-repressing factor (NFKB-repressing factor) (Transcription factor NRF) (ITBA4 protein).
DE factor NRF (ITBA4 protein).
GN NRF OR ITBA4
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=20031652; PubMed=10562553;
RA Nourbakhsh M., Hauser H.;
RT "Constitutive silencing of IFN-beta promoter is mediated by NRF (NF-

632
-435 21 192 = 0.106
89.40%

DR Pfam; PF05735, TSPC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat.
 FT NON TFR 1
 FT DOMAIN <1 14 TSP TYPE-3 1.
 FT DOMAIN 15 37 TSP TYPE-3 2.
 FT DOMAIN 38 73 TSP TYPE-3 3.
 FT DOMAIN 74 96 TSP TYPE-3 4.
 FT DOMAIN 97 134 TSP TYPE-3 5.
 FT DOMAIN 135 170 TSP TYPE-3 6.
 FT DOMAIN 171 206 TSP TYPE-3 7.
 FT DOMAIN 207 439 C-TERMINAL.
 FT CARBOHYD 424 424 N-LINKED (GLCNAC... (POTENTIAL).
 SQ SEQUENCE 439 AA; 48962 MW; 9CE8BCDD0D98070 CRC64;
 Query Match 3.1%; Score 110.5; DB 1; Length 439;
 Best Local Similarity 24.6%; Pred. No. 1.4;
 Matches 52; Conservative 19; Mismatches 89; Indels 51; Gaps 10;
 QY 112 BEKHTEGHLQCTNSDDALDPTAPNRDNNVAFSRDAARRERDGTGTCVQITNGET 171
 DB 139 QDSHDHGGQ---DACDDDD--NDGVPSDRNCRVLPNGQEDMDRDGVDGACQ---GDF 190
 QY 172 DLATMFHKLPHDLGQVTAADDFAILEDCLN--GDPSI-----CEDVPA--GDP 217
 DB 191 DADKVDKIDVCENAEVITLDFRAFQTVLDPEGDAQIDPNVVLNQGHEIVQTMNSDP 250
 QY 218 AGRLVNPTAAFAIDISGPAFSATTIPVPTLSSPELAA-----QLAEL 260
 DB 251 -GLCVGYTAENGVDGEGFFH-----VNTATDDYAGTFPGYHSSSFYVMWKQEQT 302
 QY 261 YWALARDVPMQVGTDEITTTAAANLAGMG 291
 DB 303 YWQA---NPFRAVAPGFIQLKAVKSTGPG 329
 RESULT 5
 GCP6 HUMAN STANDARD; PRT; 1819 AA.
 ID GCP6 HUMAN Q9BY91; Q9UGX3; Q9UGX4;
 AC Q96RT7; Q9BY91; Q9UGX3; Q9UGX4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Gamma-tubulin complex component 6 (GCP-6).
 GN TUBGCP6 OR GCP6 OR KIAA1669.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
 RX MEDLINE=21551508; PubMed=11694571;
 RA Murphy S.M., Preble A.M., Patel U.K., O'Connell K.L., Dias D.P.,
 RA Moritz M., Agard D., Stults J.T., Stearns T.,
 RT "GCP5 and GCP6: two new members of the human gamma-tubulin complex.";
 RL Mol. Biol. Cell 12:3340-3352(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Coile C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,

RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuayama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozeraky P., Rohlfing T.,
 RA Scheet P., Walter C., Wansley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumaneki J.P.,
 RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tiliahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [3]
 RP SEQUENCE OF 33-1819 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=21156230; PubMed=11258795;
 RA Hirosewa M., Nagase T., Murahashi Y., Kikuno R., Ohara O.;
 RT "Identification of novel transcribed sequences on human chromosome 22
 by expressed sequence tag mapping.";
 RL DNA Res. 8:1-9(2001).
 CC -|- FUNCTION: Gamma-tubulin complex is necessary for microtubule
 nucleation at the centrosome.
 CC -|- SUBUNIT: Gamma-tubulin complex is composed of gamma-tubulin, GCP2,
 CC GCP3, GCP4, GCP5 and GCP6.
 CC -|- SUBCELLULAR LOCATION: Centrosome.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q96RT7-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q96RT7-2; Sequence=VSP_001624;
 CC Note=No experimental confirmation available;
 CC -|- SIMILARITY: Belongs to the GCP family.
 CC -|- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
 CC in positions 1371 and 1758.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AF272887; AKR2968.1; -
 CC EMBL; AL022328; CAB63046.1; ALT_SEQ.
 CC EMBL; AL022328; CAB63047.1; ALT_SEQ.
 CC EMBL; AB051456; BAB33339.1; ALT_FRAME.
 CC Genew; HGNC:18127; TUBGCP6.
 CC GO; GO:0008274; C-gamma-tubulin ring complex; IDA.
 CC GO; GO:0008017; F-microtubule binding; IDA.
 CC GO; GO:0007020; P-microtubule nucleation; IDA.
 CC InterPro; IPR007259; SPC97_Spc98.
 CC Pfam; PF04130; SPC97_Spc98; 1.
 CC Microtubule; Repeat; Alternative splicing.
 FT DOMAIN 1027 1269 9 x 27 AA TANDEM REPEATS.
 FT REPEAT 1027 1053 1.

```
FT REPEAT 1054 1080 2.
FT REPEAT 1081 1107 3.
FT REPEAT 1108 1134 4.
FT REPEAT 1135 1161 5.
FT REPEAT 1162 1188 6.
FT REPEAT 1189 1215 7.
FT REPEAT 1216 1242 8.
FT REPEAT 1243 1269 9.
FT VARSPLIC 1724 1757
/FTID=VSP_001624.
FT CONFLICT 567 567 S -> L (IN REF. 3).
FT CONFLICT 1377 1377 A -> T (IN REF. 2 AND 3).
FT CONFLICT 1621 1621 L -> V (IN REF. 2).
SQ SEQUENCE 1819 AA; 200455 MW; 923576544D34594A CRC64;

Query Match 3.0%; Score 107; DB 1; Length 1819;
Best Local Similarity 24.0%; Pred. No. 19;
Matches 63; Conservative 33; Mismatches 96; Indels 70; Gaps 13;

QY 334 SPTIDAITVEPKQTFAPDLNYWDF-----DEMLNIQNGGPPAGPEELDEELRINARD 389
D 681 SRVLSALSDQMSERMALDARKQFORLKEQFVKDQERRQAARQEELDDDFSTARELRD 740
QY 390 LAR--VGFVNINTEAYRGSLLLELGAFSRPGINGPFIDSDRQAGFVNFGTSHYFRLIG 447
D 741 RERLXLSLELERKA-RQALV-----DHYSKL-- 767
QY 448 AELAQRASCYQK--WVHRFARPEALGGTLHNTIAGLDADPDISLEND-----ELLKR 501
D 768 SAEARRE--QKALMRIQ-----HRLSARLR-----FLLEDEKHIQEMLKA 808
QY 502 VAEINAAQNNEVYLLPQAIQVGSPTSPSGHATQCAFATLVKALIGLDRGCECF 561
D 809 VSEAHQPQEPDVLVSHVPQVTSFG-PEHPE--GGQCDGSGARQHSFPAWGNRPGLLT 865
QY 562 PNPVFP---SDGLLELINFEGA 580
D 866 PQLKPLAVGAGGRGLQQAEGA 887
```

RESULT 6

```
TOPI_MYCLE
ID TOPI_MYCLE STANDARD; PRT; 947 AA.
AC O69548;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Untwisting enzyme) (Swivelase).
GN TOPA OR MLO200 OR MLCB2548.31C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 408:1007-1011(2001).
CC -!- FUNCTION: The reaction catalyzed by topoisomerases leads to the
CC conversion of one topological isomer of DNA to another.
CC -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
```

```
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- MISCELLANEOUS: When a topoisomerase transiently breaks a DNA
CC backbone bond, it simultaneously forms a protein-DNA link, in
CC which a tyrosyl oxygen in the enzyme is joined to a DNA phosphorus
CC at one end of the enzyme-severed DNA strand.
CC -!- SIMILARITY: Belongs to the prokaryotic type I/III topoisomerase
CC family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AL023093; CAA18818.1; -.
CC EMBL; AL583917; CAC29708.1; -.
CC PIR; H86933; H86933.
CC HSP; P06612; IECL.
CC Leproma; MLO200; -.
CC InterPro; IPR005733; DNA_topI_bact.
CC InterPro; IPR000380; DNA_topoisomerase.
CC InterPro; IPR003601; DNATopi_ATP_bind.
CC InterPro; IPR003602; DNATopi_DNA_bind.
CC InterPro; IPR006171; Toprim_dom.
CC InterPro; IPR006154; Toprim_sub.
CC Pfam; PF01131; Topoisom_bac; 1.
CC Pfam; PF01751; Toprim; 1.
CC PRINTS; PR00417; PRTPISMRASEI.
CC SMART; SM00437; TOP1AC; 1.
CC SMART; SM00436; TOP1BC; 1.
CC SMART; SM00493; TOPRIM; 1.
CC TIGRfam; TIGR01051; copA_bact; 1.
CC TIGRfam; TIGR01051; TOPOISOMERASE_I_PROK; 1.
CC PROSITE; PS00396; TOPOISOMERASE; Complete proteome.
KW Isomerase; Topoisomerase; DNA-binding; Complete proteome.
FT ACT SITE 343 343 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 947 AA; 104055 MW; 4ACBA00C783D23AF CRC64;
```

Query Match 3.0%; Score 106.5; DB 1; Length 947;

Best Local Similarity 17.8%; Pred. No. 8.2;

Matches 120; Conservative 71; Mismatches 213; Indels 279; Gaps 26;

```
QY 27 QRAQWRRQLEGEKSLGFHPS--ETPYIKYLEGSEWKKVKLPDGTGASIKILGMARV 84
D 314 QQAGKRLRFSARTWSIAQLRYENGVIYTWRTDST---TLSESAINAR----- 360
QY 85 RIATALAVILAPCLAFDEVTASGVFPPEHHKHTGEGHLOTCNSTDALDPTAP----- 138
D 361 -----TQARQLYGEYVSDSPROYTRKVKNAQEAHAIRPAGETFA 401
QY 139 -----NRDNVAF-----ASRRDAARRERDGTG----- 161
D 402 TPDVAVCNELGDGFRIYELVWRRVTASQMAADRGTTLSLRIEGRAGEQHVVFASGRTLT 461
QY 162 -----TVQOITNGETDLATMFKHSLPHDELGOVTADDFAILDCILNGDFSICED 211
D 462 FPGFLKAYVETDELAGEADDA---ERRPLHITSQGLL-----D 498
QY 212 VPAGDPAGRLVNPATAFA-----IDISGPAFSATTIPVPTLSSPELAQAELLY 261
D 499 VIDLTPDGHATNPAPRYATEASLVKALELIGRPSVTYSSIIKTQVDRGYVQKGSALVPS 558
QY 262 WMALA-----RDVPFMQYGTDEITTAAN-----LA 288
D 559 WVAFAVFTGLLEQHFGRVLVDYDFTAAWDELDDELTAANGNEORTNMLNPNFYFGNHGVSQIA 618
QY 289 GMGGFPNLDVSI-GSDGTVDVDFPSOLF-----BATFVGV-ETGPFVSQLLV----- 332
D 619 RSGGLKKLVGNLEGIDAREVNSIKLFDDEHGRPVYVYVVKTPGYLELLAGDDGELTPQ 678
QY 333 -----NSFTTIDAITVEPKQETFA-PDLNMYVDFDEWLNIQNGGPPAGPEELDEELRIN 386
```

```
Db 679 RANLNGTLTDPDELTEVAELFATPHGRVLGVD-----PETGHEIVAKDGRY--- 726
Qy 387 ARDLARVSFVDNINTEAYRGSLLLELGAFSRPGINGPFIDS-----DROAGFVNFCTSH 441
Db 727 -----GPHYTEVLPHKDDDYGAADQTK- 749
Qy 442 YFRLIGAAELAQASCYQKQVHRFARPEALGGTTLHNTIAGDLDADPDISLLENDELLKR 501
Db 750 -----KTKGRASASQG-----PKPR-----TGSLLRSMDLQTITLEDALKL 787
Qy 502 VAEINAAQNPNVYLLPOAIOVGSPTHSPSGHATONGCAPATVLKALIGLDGRGECF 561
Db 788 L-----LPRVVG-----DPASGEITAGNGRYGPLYK-----RGKDSR 822
Qy 562 PNPVFPSSDDGLEINPEGACILITY 584
Db 823 P---LTTEDQMFITLDEALKIY 842

RESULT 7
ID DPOL RHILE STANDARD; PRF; 1016 AA.
AC Q9S1G2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase I (EC 2.7.7.7) (POL I).
OS POLA.
GN Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=384;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=9272512; PubMed=10341077;
RA Huang Y.-P., Downie J.A., Ito J.;
RT "Primary structure of the DNA polymerase I gene of an alpha-
RT Proteobacterium, Rhizobium leguminosarum, and comparison with other
RT family A DNA polymerases."
RL Curr. Microbiol. 38:355-359(1999).
CC -1- FUNCTION: In addition to polymerase activity, this DNA polymerase
CC exhibits 3' to 5' and 5' to 3' exonuclease activity (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- SUBUNIT: Single-chain monomer with multiple functions.
CC -1- SIMILARITY: Belongs to the DNA polymerase type-A family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
```

```
DR SMART; SM00479; EXOIII; 1.
DR SMART; SM00278; HHH1; 1.
DR SMART; SM00279; HHH2; 1.
DR SMART; SM00482; POLAC; 1.
DR TIGRfam; TIGR00593; pola; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydroxylase; Exonuclease; DNA-binding.
FT DOMAIN 1 308 5'-3' EXONUCLEASE (BY SIMILARITY).
FT DOMAIN 394 630 3'-5' EXONUCLEASE (BY SIMILARITY).
FT DOMAIN 768 1016 POLYMERASE (BY SIMILARITY).
SQ SEQUENCE 1016 AA; 111491 MW; 35738F85B8C70514 CRC64;

Query Match 3.0%; Score 106.5; DB 1; Length 1016;
Best Local Similarity 18.8%; Pred. No. 9.1;
Matches 110; Conservative 86; Mismatches 215; Indels 174; Gaps 26;

Qy 177 FHKSPLHDELQVTD-----DFAILEDICILN-----GDFSICEDVPAGDPAGRLVNPATAFA 229
Db 295 FLXTMEFTTLTRRVAEACDCDAIEPAIVRIEWGETARGPDLDAAEP-----EPVAGGI 349
Qy 230 IDISGPAFSATTIPP-----VPTLSSP-ELAAQLAELV----- 261
Db 350 PEVSG---ESVNPMPRAKSAVEGAFSPADLAKARAEAFATLPFDHSAYTIRDLVTLD 406
Qy 262 -MMALARDVPFMYGTDEITTTAAANLAGMGFPNLDVAISGDSGTVDYDFPSQLFRATFVG 320
Db 407 RWIADARATGLVAPDTE--TTSLDAMQALVGF---SLAI-ADNTADPTGKIRAAVVP 459
Qy 321 VETGPFVSQLVNSFTIDAITVEPKQETPAPDLNVMVDFDEWLNIQNGGPPAGPEEDEE 380
Db 460 LVHKNVGVDLLGGGLADNQI---PMRDAL-PRLKALLEDESVLKVA-----QN 503
Qy 381 LRF-----IRNARDLARVSFVDNINTEAY-----RGSLLLELGAFSR 418
Db 504 LKDYDLLLKKYGIETRSFDDTMLSIVLDAGTAGHMDPLSEKFLGHTPTPYKDVAGSGK 563
Qy 419 PGINGPPIDSDRQAGFVNFCTSHYFRL-----IGAAELAQASCYQKQW---VHRPAR 468
Db 564 ANVTFDLVDIDRATHVAEADADVTLLRLWLKPLRAAAGL---TSVYERLERPLLPVLAR 620
Qy 469 PEALGTLHNTIAGDLDADPDISLLENDELLKRVAEINAAQNPNVYLLPOAIOVQSP 528
Db 621 MEARGITVDRIQLSRL-----SGELAAQGAARL-----EDEIYVLAGERFNIGSP 664
Qy 529 TH-----PSYPSGHATONGAFATVLKALIGLDRCG-----E 559
Db 665 KQIGDILFGKMGSLGSGSKTKTGWSTSAQVLEDLAAAGFELPKIVDWRQVTKLKSTYTD 724
Qy 560 CFNPVPFPSSDDGLE-----LINFEAGACTYEGEINKLVNVAFGQRLMGLHYRFDGIQGL 614
Db 725 ALPGVYHPETKRVHTSVSLASTTTGRLSSSEPNLQNPVTAEGRKI----- 771
Qy 615 LIGETTIVRTLHQLMTFAEATPEPRLFT-----GEVILKFDQG 654
Db 772 ---RTAFISTPGHKLIS-ADYSQIELRVLAHVAEIPOLTKAFEDG 812

RESULT 8
LPXC PSESIM STANDARD; PRF; 303 AA.
AC Q87WZ1;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase
DE (EC 3.5.1.-) (UDP-3-O-acetyl-GlcNAc deacetylase).
GN LPXC OR PSPTO4402.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=22834015; PubMed=12928499;
RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA Winn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA Nelson W.C., Davidse T., Zafar N., Zhou L., Liu J., Yuan O.,
RA Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
RA Alfano J.R., Cartinour S., Chatterjee A.K., Delaney T.P.,
RA Lazarowitz S.G., Martin G.B., Schaeffer D.J., Tang X., Bender C.L.,
RA White O., Fraser C.M., Collier A.;
RT "The complete genome sequence of the Arabidopsis and tomato pathogen
RT Pseudomonas syringae pv. tomato DC3000.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC -1- FUNCTION: Involved in the biosynthesis of lipid A, a
CC phosphorylated glycolipid that anchors the lipopolysaccharide to
CC the outer membrane of the cell (By similarity).
CC -1- CATALYTIC ACTIVITY: UDP-3-O-(3-hydroxytetradecanoyl)-N-
CC acetylglucosamine + H(2)O = UDP-3-O-(3-hydroxytetradecanoyl)-
CC glucosamine + acetate.
CC -1- PATHWAY: Lipid A biosynthesis; second step.
CC -1- SIMILARITY: Belongs to the lpxC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016871; AA057851.1; -.
DR HAMAP; MF_00388; -.
DR InterPro; IPR008969; CarboxypepD_reg.
DR IPR004463; LpxC.
DR Pfam; PF03331; LpxC; 1.
DR Hydrolase; lipid A biosynthesis; Lipid synthesis; Complete proteome.
SQ SEQUENCE 303 AA; 33227 MW; D72873E8A49C6434 CRC64;

Query Match 3.0%; Score 106; DB 1; Length 303;
Best Local Similarity 25.4%; Pred. No. 1.7;
Matches 76; Conservative 39; Mismatches 106; Indels 78; Gaps 19;

QY 307 VDPSQL-FRATFVGTGPFVSVQLVNSFTDAITVEPKQETPAP---DLNVMVDFEW 362
Dy :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 44 LDPVQIPARAENVGDTT--LSTLVNG-DVKVDTVEHLLSAMAGLIDNAYVELSASE 99

QY 363 LNQLNG--GP-----PAGPEELDEELRFTNRADLA-----RVSVVDNINTEAVRGL 408
Dy :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 100 VPIMDSAGPFPVLIQAGLEEQAPKKFIRLREVTVVEGGKRTFV-----PPGFK 153

QY 409 ILLELGAFSRPGINGPPIIDSPR-QAGFVNFQTSYHFRLLGAAELAQASCYQKQVHRFA 467
Dy :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 154 VSEID-FDHP-----VFRDRTQASVDFSSITSFVKEVSRRFTFGFMSDIEYLRKLN- 205

QY 468 RPEALGTLNNTIAGLDADAF-DISLLENDELLKR-----VAEINAAQNPNVEYLLPQ 521
Dy ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 206 ---ALGGSVENAIIVVDKGVNLNEDGLRYEDEFVKHKILDAIGDL-----YLLGN 251

QY 522 AIQVGSPTSPSYSGHATONGAFATVLKALIGLDGCGCFPNFVPSDDGLLEINFEGA 580
Dy :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 252 SL-IGF--FRGFKSGHALNN-----RLRLALI-----EQTDWVVTTFEDA 289

RESULT 9
HISX METH
ID HISX METH STANDARD; PRT; 426 AA.
AC 02632;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Histidinol dehydrogenase (EC 1.1.1.23) (HDH).
GN HISD OR MTH225.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldridge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- FUNCTION: Catalyzes the sequential NAD-dependent oxidations of L-
CC histidinol to L-histidinolaldehyde and then to L-histidine (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: L-histidinol + 2 NAD(+) + H(2)O = L-histidine
CC + 2 NADH.
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- PATHWAY: Histidine biosynthesis; ninth (last) step.
CC -1- SIMILARITY: Belongs to the histidinol dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000809; AAB84731.1; -.
DR PIR; H69127; H69127.
DR HAMAP; MF_01024; -.
DR InterPro; IPR001692; Histidinol_dh.
DR Pfam; PF00815; Histidinol_dh; 1.
DR PRINTS; PR00083; HOLDHGRNAS.
DR ProDom; PD002680; Histidinol_dh; 1.
DR TIGRFAMs; TIGR00069; hisd_1.
DR PROSITE; PS00611; HISOL_DEHYDROGENASE; 1.
DR Histidine biosynthesis; Oxidoreductase; NAD; Metal-binding; Zinc;
KW Complete proteome.
FT ACT_SITE 323 323 BY SIMILARITY.
FT ACT_SITE 324 324 BY SIMILARITY.
FT METAL 255 255 ZINC (BY SIMILARITY).
FT METAL 258 258 ZINC (BY SIMILARITY).
FT METAL 357 357 ZINC (BY SIMILARITY).
FT METAL 416 416 ZINC (BY SIMILARITY).
SQ SEQUENCE 426 AA; 46419 MW; A9A0F2AA25B8594A CRC64;

Query Match 3.0%; Score 106; DB 1; Length 426;
Best Local Similarity 21.8%; Pred. No. 2.8;
Matches 88; Conservative 42; Mismatches 136; Indels 136; Gaps 16;

QY 130 DDALDPTAPNRDNVAFASRRDAARRDGTGTCVQITNGETLATMFKHSLPHDELQV 189
Dy ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 25 DDVLGPVA-----DIIISVROGDDALRELTCGRFDGVTVENFVSREEIEA 71

QY 190 TADDPFAILEDCTINGDFPSICE-----DVPAGDPAGRLVNPATAFADISG--- 234
Dy :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 72 HKNLEPGVKEALREAAASNIIEFHRMQMPSGWMSEVRPGVMAGQLVRPDSVGCYIPGGA 131

QY 235 --PAFSATTI-----PPVPTLSPE---LAAQLA---ELVWMLARDVPF 271
Dy ||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 132 VYPTILMTVIPARIAGVERIVCTPPAQDGSVPDVAVLVADWAGASEIRVGGQAQVAA 191

QY 272 MOYGTD-----EITTTAAANLAGMGFPNLD-----AVSIGSDGTVDPPSQL 313
```


FT	DOMAIN	2595'	2652	SUSHI 16.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2657	2717	SUSHI 17.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2722	2775	SUSHI 18.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2780	2833	SUSHI 19.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2838	2891	SUSHI 20.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2896	2953	SUSHI 21.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	2958	3011	SUSHI 22.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3019	3072	SUSHI 23.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3077	3131	SUSHI 24.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3136	3191	SUSHI 25.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3196	3249	SUSHI 26.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3254	3307	SUSHI 27.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3315	3369	SUSHI 28.	(GLCNAC. .)	(POTENTIAL).
FT	DOMAIN	3374	3429	SUSHI 29.	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	36	36	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	53	53	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	324	324	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	372	372	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	687	687	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	786	786	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	929	929	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1055	1055	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1089	1089	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1134	1134	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1243	1243	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1499	1499	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1554	1554	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1672	1672	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1744	1744	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1892	1892	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1982	1982	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2118	2118	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2249	2249	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2254	2254	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2287	2287	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2458	2458	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2500	2500	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2674	2674	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2705	2705	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2825	2825	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2895	2895	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2915	2915	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3062	3062	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3121	3121	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3130	3130	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3157	3157	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3171	3171	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3181	3181	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3239	3239	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3327	3327	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3485	3485	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3492	3492	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3575	3575	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3581	3581	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	3590	3590	N-LINKED	(GLCNAC. .)	(POTENTIAL).
FT	VARSPLIC	1	22	MGKSGFTFWNLVFLTVSCVK -> MWSWFLCWKPQLDR		
FT				QTAS (in isoform 2).		
FT				/FTId=VSP_009047.		
FT	VARSPLIC	307	410	Missing (in isoform 3).		
FT				/FTId=VSP_009048.		
FT	VARSPLIC	2655	2719	Missing (in isoform 3).		
FT				/FTId=VSP_009049.		
FT	VARSPLIC	1762	1831	Missing (in isoform 4).		
FT				/FTId=VSP_009050.		
FT	VARSPLIC	2592	2638	AISCGIPKPAFTNGGILFTDYLGVTRVYFCNDGVRLSSKEL		
FT				TTAVCQ -> GEVYAKGNKMNVRLAPFNWITFNSEN		
FT				GNIRKHIVNSPHKKA (in isoform 5).		
FT				/FTId=VSP_009051.		
FT	VARSPLIC	2639	3670	Missing (in isoform 5).		
FT				/FTId=VSP_009052.		
FT	VARIANT	182	182	I -> M (in dbSNP:2219898).		
FT				/FTId=VAR_017404.		

Query Match		3.0%	Score 106;	DB 1;	Length 3670;
Best Local Similarity		19.0%	Pred. No. 63;		
Matches 142;		Conservative 92;	Mismatches 253;	Indels 262;	Gaps 40;
Qy	2	LCHAADTRGSPMPDTCVLLTSEQRKAKWRQLEGEKSLGPHSPSTPYI	-----	52	
Db	197	ICRAEDACGTMRGSGI---ISSPSFPNHNADCTWIVABPGDTISLIFTDFQWEE	253		
Qy	53	KY---LEGSE-----TWKVKVLPD-----GISA-----	73		
Db	254	KDYLEIEGSEPTIWSGMNIPPIISKNKWLHFVTDNSHRYRGSFAPYQSSILTH	313		
Qy	74	SKILGIMARVRIATALLAVVLAAPCLAFDEVTSAGVPEEHEKHTGEGRHLQTCNSDAL	133		
Db	314	TTSTGELEENRITTTGAIVASTPA---DVTVSSVTAVTIHRLSEQRQVQVTLRNSGL	369		
Qy	134	DPT-----APNRDNNVAFASR-RDAARRRDGTGTVCQITN---GETDLATMFKHSLPH	183		
Db	370	DPNTSKDGLSPHPADTQSTRRRPRHABQIER--TKELAVVTVHRVKKAIKFKSRGFKLP-	426		
Qy	184	DELGOVTADDPFAILEDCLNGDFSIQEDVPAGDPA-GRLV-----NPTAAFAIDISGPA	236		
Db	427	---GKONSNKFSILNEGKITASNLCPD---PGEPEGKRIQSGDFSLGTVQFSCDEYVL	481		
Qy	237	FSATTIP-----PVPTLSSPELAAQL---AELYWMA	264		
Db	482	QGAKSITCQRIAEVFAAWSDRHPVKVKTCGSLNQLQSPSGTFTSPNPPFQYDSNAQCVWVI	541		
Qy	265	LARDV-PFMOYGTDEITTTAAANLAGMGPFLNDAVIGSDGTV-DPFSQLFRATFVGVE	322		
Db	542	TAVNTNKVIQINFEEDLEI-----GYDTLTIGDGEVGDG-----RTVLQVL	584		
Qy	323	TGPEVSQLVNSFTTIDAITVEPKQETPAPDLNVMVDPDEWLNQNGGPPAGPELDEBLR	382		
Db	585	TGSFV-----PDIVSMSSQMLHLQT-----DESVCVG	614		
Qy	383	FIRNARDLARVSVDNINTEAYRGSLLLELGAFRSPGNGPFIDSDRQAGFVN-----F	437		
Db	615	FKVNYKEIKESCGD-----PQT-PLYGIREGDFGFSNRDLRF	651		
Qy	438	GTSHYFLIGAAELAQRASYQKQVH-----RPAR-----PEALGGTILH	477		
Db	652	ECQFGPELIGEKSIQCOEN--NQWSANIPICI PCLSNFTAPMGTVLSDPYPEGYGNLN	709		
Qy	478	N--TIAGDLQADPDLSLENDLLELLKRVAEINAQNPNNVEVYLLPOAIOVGSPTHPSPS	535		
Db	710	CIWTIISDPCSRHLSP--NDFLESQDFLAVKDGDS-----PESPILGTTFTGAEVPS	761		
Qy	536	GHATONGAFATVILKALIGLD-----RG-----GECFPNPVPSD-----DGLLEL-	574		
Db	762	-HLTSN---SHILRLBFOADHSMGRGFNITYNTFGHNEC-PDQGIPIINARRFGDNFQLG	816		
Qy	575	-----INPEG-----ACLYTEGEI	588		
Db	817	SSISVICEEGFIKTQGTETITCILMDGKV	845		
RESULT 11					
SECD_RHOCA					
ID	SECD_RHOCA	STANDARD;	PRT;	554 AA.	
AC	O33517;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	Protein-export membrane protein secd.				
GN	SECD.				
OS	Rhodobacter capsulatus (Rhodopseudomonas capsulata).				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;				
OC	Rhodobacteraceae; Rhodobacter.				
OX	NCBI_TaxID=1061;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SB1003 / St Louis;				


```
Db 476 PAFTATSSSQIKQIGPELPRP-----NPVRIPTRRRELASFGIKLP 518
Qy 152 AAR-----RRDGTGTVCQITNGETDLATMFHKSPLHDELQGVTDADFAILEDICILNGD 205
Db 519 SQMAEQELRERDQ-----DETQNPQWAASSYG-----TEITSDEDAALQAILRKA 565
Qy 206 PSICEDVPAGDPAGRLVNPFAAFIDIS--GPAFSATTIPVPTLSPELAAQLAELVYM 263
Db 566 F-----ADQOSRYALSTLAEOSSITERSPAAEMPTTTSQVSDLEDEQALQAEALRQA 618
Qy 264 ALARDVPMQVGTDEITTTAAANLAGMG--GFNLDVAVSGISGTVDPSPQLPRAFTVG 320
Db 619 FAQQQ--HRYGAGTGTNDVNDVIRSVDTSTFTSPIADLVDSPREPUFTL----- 669
Qy 321 VETGPFVSQLVNSFTIDAITVEPKQETFAPD 352
Db 670 ---SPYDEITVD---EPVQLECKERSLLQD 694

RESULT 14
COMP_MOUSE
ID_COMP_MOUSE STANDARD; PRT; 755 AA.
AC QROG6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cartilage oligomeric matrix protein precursor (COMP).
GN COMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=20505681; PubMed=11052496;
RA Fang C., Carlson C.S., Leslie M.P., Tulli H., Stoleran E., Perris R.,
RA Ni L., Di Cesare P.E.;
RT "Molecular cloning, sequencing, and tissue and developmental
RT expression of mouse cartilage oligomeric matrix protein (COMP).";
RL J. Orthop. Res. 18:593-603(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 28-71.
RX MEDLINE=22313308; PubMed=12426368;
RA Ozbek S., Engel J., Stetefeld J.;
RT "Storage function of cartilage oligomeric matrix protein: the crystal
RT structure of the coiled-coil domain in complex with vitamin D(3).";
RL EMBO J. 21:5960-5968(2002).
CC -!- SUBUNIT: Pentamer; disulfide-linked (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the thrombospondin family.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TSP type-3 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF033530; AAD01972.1; -.
CC PDB; 1M29; 28-JAN-03.
CC MGD; MGI:88469; Comp.
CC InterPro; IPR002048; EF-hand.
CC InterPro; IPR001881; EGF Ca.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR003367; tsp_3.
CC InterPro; IPR008859; TSPC.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF05735; TSPC; 1.
CC Pfam; PF02412; tsp_3; 13.
CC
```

```
DR SMART; SM00179; EGF CA; 2.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 3.
DR PROSITE; PS01187; EGF_CA; 2.
KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;
KW Signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 755
FT DOMAIN 21 84
FT DOMAIN 85 124
FT DOMAIN 125 177
FT DOMAIN 178 220
FT DOMAIN 223 265
FT DOMAIN 295 330
FT DOMAIN 331 353
FT DOMAIN 354 389
FT DOMAIN 390 412
FT DOMAIN 413 450
FT DOMAIN 451 486
FT DOMAIN 487 522
FT DOMAIN 523 755
FT DISULFID 68 68
FT DISULFID 71 71
FT DISULFID 89 100
FT DISULFID 94 109
FT DISULFID 112 123
FT DISULFID 129 140
FT DISULFID 134 149
FT DISULFID 152 176
FT DISULFID 182 195
FT DISULFID 189 204
FT DISULFID 207 219
FT DISULFID 227 241
FT DISULFID 235 251
FT DISULFID 253 264
FT DISULFID 280 285
FT DISULFID 290 310
FT DISULFID 326 346
FT DISULFID 349 369
FT DISULFID 385 405
FT DISULFID 408 428
FT DISULFID 446 466
FT DISULFID 482 502
FT DISULFID 518 739
FT CARBOHYD 119 119
FT CARBOHYD 740 740
SQ SEQUENCE 755 AA; 82352 MW; 7DDFCF443589A0B7 CRC64;

Query Match 2.9%; Score 103.5; DB 1; Length 755;
Best Local Similarity 24.2%; Pred. No. 9.9;
Matches 51; Conservative 19; Mismatches 90; Indels 51; Gaps 10;

Qy 112 EEKHGTGEGRHLQCTNSDDALDPTAPNRNNDVAFASRRDAARRERDGTGVCQITNGET 171
Db 455 QDSHDCKG--DACDDDD--NDGVDSRDNCLVNPQGEDNDRDGVGDAQC---GDF 506
Qy 172 DLATMFHKSPLHDELQGVTDADFAILEDICILN--GDFS-----CEDVFA--GDP 217
Db 507 DADKVIDKIDVCPENAEVTLTDFRAFQTVVLDPEGDAQIDPNWVVLNQGMIEIVQTMSNDP 566
Qy 218 AGRLVNPTAAFAIDISGPAFSATTIPVPTLSPELAA-----QLAEL 260
Db 567 GLAVGYTAFNGVDPEGTFFH-----VNTATDDYAGPIFGYQDSSSFYVVMKQMEQT 618
Qy 261 YWMALARDVPFMQVGTDEITTTAAANLAGMG 291
Db 619 YWQA----NPFRAVAEPGIQLKAVKSTGPG 645

RESULT 15
COMP_RAT
```

